



Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

VOL. 72
1958

Bureau of Educ. & Psyl. Res.

EX-111

9.3.81

Acc 57.38

NORMAN L. MUNN, EDITOR

*Bowdoin College
Brunswick, Maine*

Consulting Editors

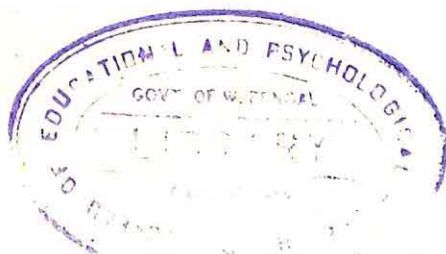
ANNE ANASTASI
FRANK A. BEACH
ARNOLD M. BINDER
W. J. BROGDEN
ROBERT R. BUSH
JOHN F. DASHIELL
JAMES J. GIBSON
D. O. HEBB
EDNA HEIDBREDER
FRANCIS W. IRWIN

JAMES J. JENKINS
HAROLD E. JONES
DANIEL KATZ
BOYD McCANDLESS
DONALD W. MACKINNON
QUINN McNEMAR
LORRIN A. RIGGS
CARL R. ROGERS
RICHARD L. SOLOMON
ROSS STAGNER

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.



CONTENTS OF VOLUME 72

Whole No.

- 454 THE METHOD OF PAIRED DIRECT AND PROJECTIVE QUESTIONNAIRES IN THE STUDY OF ATTITUDE STRUCTURE AND SOCIALIZATION. J. W. Getzels and J. J. Walsh. Pp. 34.
- 455 THE USE OF ROTTER'S SOCIAL LEARNING THEORY IN DEVELOPING A PERSONALITY INVENTORY. Shephard Liverant. Pp. 23.
- 456 PERSONALITY OUTCOMES OF CLIENT-CENTERED CHILD THERAPY. Elaine Dorfman. Pp. 22.
- 457 THE CONCURRENT VALIDITY OF SIX PERSONALITY AND ADJUSTMENT TESTS FOR CHILDREN. Louis M. Smith. Pp. 30.
- 458 A THEORETICAL AND EXPERIMENTAL CONSIDERATION OF THE RORSCHACH HUMAN MOVEMENT RESPONSE. Gerald F. King. Pp. 23.
- 459 CONSISTENCY OF PROJECTIVE MOVEMENT RESPONSES. Joseph F. Okarski. Pp. 26.
- 460 PATTERNS OF MENTAL FUNCTIONING ASSOCIATED WITH PREJUDICE IN CHILDREN. Bernard Kutner. Pp. 48.
- 461 CONTROL OF BEHAVIOR IN CHIMPANZES AND PIGEONS BY TIME OUT FROM POSITIVE REINFORCEMENT. Charles B. Ferster. Pp. 38.
- 462 RIGIDITY-FLEXIBILITY AND INTELLIGENCE: A CROSS-SECTIONAL STUDY OF THE ADULT LIFE SPAN FROM 20 TO 70 YEARS. K. Warner Schaie. Pp. 26.
- 463 ATTITUDE CHANGE: INSTABILITY OF RESPONSE AND ACQUISITION OF EXPERIENCE. George Katona. Pp. 38.
- 464 COMPARISON OF DIRECT AND PROJECTIVE METHODS OF PERSONALITY ASSESSMENT UNDER DIFFERENT CONDITIONS OF MOTIVATION. Anthony Davids and Henry Pildner, Jr. Pp. 30.
- 465 CEREBRAL LOCALIZATION OF INTELLECTUAL PROCESSES. Gerald H. Shure and Ward C. Halstead. Pp. 40.
- 466 FACTOR ANALYTIC STUDY OF VISUAL MEMORY. Raymond E. Christal. Pp. 24.
- 467 THE EFFECT OF A DISTURBING SITUATION UPON THE STABILITY OF VARIOUS PROJECTIVE TESTS. Rhoda Lee Fisher. Pp. 23.
- 468 TWO PROCESSES OF CONCEPT FORMATION. Harriett Amster Podell. Pp. 20.
- 469 EXPRESSIVE GRAPHIC MOVEMENTS AND THEIR RELATIONSHIP TO TEMPERAMENT FACTORS. Max Talmadge. Pp. 30.
- 470 AGE-SEX ROLES AND PERSONALITY IN MIDDLE AGE: A THEMATIC APPERCEPTION STUDY. Bernice L. Neugarten and David L. Gutmann. Pp. 33.

Psychological Monographs: General and Applied

The Method of Paired Direct and Projective Questionnaires in the Study of Attitude Structure and Socialization¹

J. W. GETZELS
University of Chicago
and J. J. WALSH
Boston College

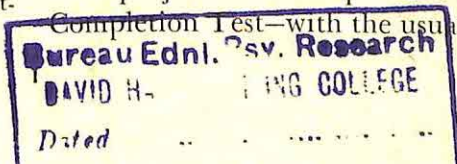
THERE is hardly a systematic position in current behavior theory that does not include some structural concept of *levels* of reaction. These levels have variously been conceived as "private-public" (14), "covert-overt" (4), "latent-manifest" (8), "subconscious-conscious" (5), "genotype-phenotype" (12), and so on. These constructs are not, of course, all of equal generality, but the very pervasiveness of *some* concept of levels points to the crucial significance of the "depth" dimension of reactions in understanding behavior. Systematic penetration into the nature of the relationship between these levels is of critical importance in the study of personality and attitudes.

Research in this area has not, however, been as fruitful as the underlying notions would seem to merit. The difficulty has been, not with the concept of an "in-

ternal-external" dimension as such, but with the lack of an efficient technique for determining the "internal" and "external" reactions to the *same stimulus* and for examining the relationship between them. On the one hand, there have been numerous instruments for eliciting the overt levels of behavior—e.g., self-report personality tests, attitude scales, interest inventories, and such; on the other, numerous instruments for eliciting the covert levels—e.g., the Rorschach, TAT, Word-Association, and such. But between the two there has been an unbridged gap. With responses to what direct procedure, for example, are reactions to the Rorschach or the TAT or the Word-Association to be compared; and, conversely, with responses to what projective instrument are replies to the Bell, the Bernreuter, or the Guilford-Martin to be compared?

What is required first of all in this area is a method for eliciting direct and projective responses to the *same* object of inquiry in such way that the responses may be immediately and objectively compared in the *same* terms. The Method of Paired Direct and Projective Questionnaires (PDPQ) was devised to meet the indicated need by combining a widely used projective technique—the Sentence Completion Test—with the usual form of

¹ The preparation of this monograph was facilitated by a grant from the Social Science Research Committee of the University of Chicago. The initial research and the first study were made possible by a grant from the Anti-Defamation League and the Social Relations Laboratory at Harvard University where the senior author held a Livingston Fellowship. The help of Gordon W. Allport, Eugenia Hanfmann, Richard L. Solomon, and Samuel A. Stouffer is gratefully acknowledged. The second study was done at the University of Chicago, and the authors are indebted to Allison Davis and the late Karl J. Holzinger for their invaluable advice and assistance.



direct questionnaire, the two instruments fulfilling the stated requirements.

The current report has two objectives:

1. To describe the rationale, construction, scoring, and interscorer reliability of the PDPQ.

2. To present the results of two studies designed to investigate the usefulness and validity of the method. The first study deals with the nature of the relationship between overt and covert levels of attitude structure as a function of *differential objects of inquiry*; the second study with the nature of the relationships between overt and covert levels of attitude structure as a function of *differential patterns of socialization* represented by the age, sex, sibling position, and social class characteristics of respondents.

THE QUESTION-ANSWER PROCESS AND THE METHOD OF PAIRED DIRECT AND PROJECTIVE QUESTIONNAIRES

The rationale upon which the method and the studies are based derives from a recent statement of a theory of perception (3), and of the question-answer process elaborated elsewhere by the senior author (7). Central to this theory is the intervening variable: the *personal hypothesis*. The hypothesis may be regarded as an internal state of readiness to react selectively to objects in the environment. The selectivity of perceiving, problem-solving, imagining is said to be governed not only by the objective stimulus but by the personal hypothesis. Answering questions, no less than perceiving and problem solving, must presuppose already established hypotheses or internal sets. If an individual is ready to answer a question at all, he is ready to answer it in a given way. The question serves as a stimulus to trip off this personal hypothesis.

But this is ~~only~~ the first step. The per-

sonal hypothesis may make itself manifest in a blush, a stammer, a block, a so-called "spontaneous response," or an "impulsive remark," but before it is ordinarily externalized in a verbalization for which the individual is willing to stand responsible, two other factors enter into consideration: the situation of the questioning and the motives of the individual in the situation. Appropriate behavior may require revealing or concealing the personal hypothesis. The *second* step in the question-answer process is the referral of the question-stimulus to the situational context and the checking of the initially aroused personal hypothesis against the requirements of the situation. As the clinician, psychiatrist, and interviewer has learned, what a person says about himself may represent a number of motives. His response to an inquiry may be prompted as much by motives for giving an inaccurate account as for giving an accurate account of his reactions. The *third* step is the formulating of a verbal response—the *expressed reaction*—to the question that will facilitate or at least not hinder the respondent's adjustment in the light of personal needs relative to situational demands. Exposing the private impulse to public view may aid or hinder the subject's adjustment, and the personal hypothesis is liable to defense and censorship—the *expressed reaction* may be a communication or a distortion of the *personal hypothesis*.

It is this relationship between personal hypothesis and expressed reaction or between the covert and overt levels of attitude that is crucial to the present method and series of studies. The experimental hypotheses for empirical test will be derived from this formulation of the question-answer process and will focus upon systematic relationships between these

levels of response, given certain groups of subjects, certain objects of inquiry, and certain situational factors.

The technique of Paired Direct and Projective Questionnaires is advanced as singularly capable of eliciting the kinds of behavior required for the study of the indicated relationships. With respect to a given object of inquiry, it provides a measure of the personal hypothesis, the expressed reaction, and the relative congruence or discrepancy between the two. The method utilizes two instruments of the sentence completion type, one of which, phrased as a self-report questionnaire, is direct in nature, and the other, containing exactly the same items phrased in the third person, is projective in nature.

In responding to the direct instrument, the subject is fully aware that what he says refers to himself, and that a value judgment within some particular frame of reference of which he is also aware may be made on the basis of his responses. The subject may respond with a deliberately invoked standard of what he considers appropriate to himself or to others in the particular setting, and he may distort or censor the original reaction tendencies in order to make his behavior appear consistent with the criteria invoked for the situation. This is not to say that the response to a direct question must inevitably be swayed more by circumstance than by accuracy, or that there will invariably be discrepancies between the manifest response revealed by the direct type instrument and the latent reaction purportedly revealed by the indirect type instrument. The direct response must not be considered a superficial response in the sense that it is not a useful datum or is not valid *at its own level of reaction*. In this study, the position is taken that the response to a direct

question represents the level of behavior at which the individual permits society to look at him—it is the *expressed reaction* of the preceding conceptualization of the question-answer process.

In responding to the projective instrument, the subject is unaware that what he says refers to himself, and that a value judgment within a particular frame of reference may be made on the basis of his responses. In the present projective instrument the subject is not required to report his own reaction to a particular stimulus, but—under the guise of a verbal speed test—to attribute to an ambiguous third person a reaction to the same stimulus. The basic assumption is that when a subject is pressed to attribute behavior to others, and he perceives the situation as devoid of personal reference, he tends to reveal levels of behavior he might otherwise be disposed to distort or censor. As Sanford has observed:

For revealing *underlying* forces the most fruitful procedures are those which require the subject to attribute actions and motives to other people. Faced with this task what comes most readily to the subject's mind are his own feelings, emotions, and attitudes when faced with situations similar to those of the characters presented to him (15, p. 257).

In effect, in responding to a projective instrument, the subject is said to provide a picture of his *personal hypotheses*, as previously defined.

If these assumptions regarding the direct and projective instruments of the PDPQ are sound, then the response to the direct instrument should in fact approximate what has been called the expressed reaction and the response to the projective instrument what has been called the personal hypothesis. The behavior of the direct and projective responses relative to each other for a given object of inquiry should approximate the postulated relationships between the per-

sonal hypotheses and expressed reactions to be derived from the model of the question-answer process for known respondents and known objects of inquiry. Empirical verification of these relationships will provide evidence for the usefulness and validity of the framework and the method, in addition to yielding substantive data regarding the intrapersonal levels of reaction for those aspects of personality and attitudes at issue in the given studies.

THE METHOD OF PAIRED DIRECT AND PROJECTIVE QUESTIONNAIRES: CONSTRUCTION OF INSTRUMENTS, SCORING PROCEDURES, AND RELIABILITY OF SCORING

Constructing the PDPQ entails choosing the objects of research interest (attitudes, behaviors, opinions), inventing situations representing the objects, and writing first and third person incomplete sentence stubs describing the situations in such way that the sentence may readily be completed by manifesting a reaction to each stimulus situation. If, for example, the object of inquiry is *leadership*, the situation might present a choice between accepting or failing to accept a request to be "in charge." The sentence stub in the third person form might be: "When they asked Frank to be in charge, he. . . ." In the first person form the corresponding item would be: "When they asked me to be in charge, I. . . ." Again, if the object of inquiry is attitudes toward the *Negro*, the third person stub might be: "If Negroes began being admitted to the club, Bill. . . ." The first person item would then read: "If Negroes began being admitted to the club, I. . . ."

The third person incomplete sentences are administered *projectively* disguised as a verbal speed test, the respondent

being unaware that the responses will be evaluated as attitudes, preferences, sentiments, and such. The response to this instrument represents the personal hypothesis of the model of the question-answer process. After an appropriate time lapse the first person incomplete sentences are administered *directly* as a personal questionnaire, the respondent being aware that the responses will be evaluated as his attitudes, preferences, sentiments, and such. The response to this instrument represents the expressed reaction of the model of the question-answer process.

The items are written so as to conform to three major requirements: First, they must be of such a nature and structure as to facilitate a dichotomous classification of the responses, usually as either positive (socially acceptable) or negative (socially unacceptable). This structuring undoubtedly restricts the range and variety of responses, but since the present studies were concerned only with quantitative comparisons, clinical richness was sacrificed for precision of scoring. Second, the items are to be free of affect-laden words that might exert undue influence on the nature of the response. Thus, words like "good," "bad," "angry," "sad,"—words that might demand uniformly positive or negative responses—are avoided. Third, the items should as far as possible present behavioral situations corresponding to areas identified in theoretical or empirical studies as having differential values for varying subsections of the population.

Although the responses to the instruments lend themselves both to qualitative and quantitative scoring, in the present studies only a quantitative treatment of the data is undertaken. The basic score unit is the negative response, and a subject's score on the projective and

on the direct instrument is the number of negative completions given on each of the instruments. The third metric is the proportion of negative responses to the projective instrument reversed to non-negative responses to the direct instrument, thus providing a measure of the discrepancy between the personal hypotheses and the expressed reactions.

Scoring keys were prepared as aids for consistent and objective scoring. It is, of course, impossible to reproduce the keys here, but several examples will serve to clarify the nature of the responses and the scoring procedures.

A typical item was: "When they asked Frank to be in charge, he. . . ." Typical responses were:

didn't want to.	accepted and tried to do
quit.	his best.
said no.	was happy.
said he couldn't.	accepted with gladness.
worried.	took over.
wasn't sure.	accepted.
did not want to be.	felt good.
refused.	agreed.
replied he was in-	did his best.
capable.	felt honored.
declined.	was pleased.
	liked the idea.

All responses in the left column were scored negative; those in the right hand column, positive.

Another item was: "Working with others all the time made him. . . ." The responses were:

nervous.	happy.
uneasy.	very happy.
feel uncomfortable.	more sociable.
mad.	content.
angry.	easy to get along with.
jumpy.	cooperative.
unhappy.	congenial.
annoyed.	pleasant.
irritable.	sociable.
upset.	more efficient.
sick.	very pleasant.
tired.	feel good.
insignificant.	very agreeable.
a little uneasy.	a very broad-minded per-
leery of his ability.	son.
	happy and enthusiastic.

Again, there was little difficulty in scoring the responses on the left as negative and those on the right as positive. The number of indeterminate responses, i.e., those that could not be scored as either negative or positive, varied with the nature of the item. For the instruments as a whole, such indeterminate responses accounted for only a very small fraction of the total number of responses. Actually, these indeterminate responses were of little consequence, since only clearly negative projective and direct responses were utilized in the analyses.

The crucial question regarding a scoring technique of the sort utilized here is the extent to which the scoring is influenced by the subjective judgment of the scorer. This is the problem of interscorer reliability. This problem was investigated by correlating the scores assigned to a random sample of 20 pairs of tests, each test containing 38 items, with the scores assigned to these tests by two other independent scorers. The matrix of these correlations is presented in Table 1. The original scorer is designated as 1, and the two scorers engaged for the reliability study are designated as 2 and 3.

Certainly there can be no dissatisfaction with these interscorer reliabilities. For the process of quantifying essentially qualitative data, these reliability coefficients reflect excellent consistency from one scorer to another. It is amply clear

TABLE 1
INTERSCORER RELIABILITIES FOR PROJECTIVE,
DIRECT, AND DISCREPANCY SCORES

Scorer	Score					
	Direct		Projective		Discrepancy	
	1	2	1	2	1	2
2	.977		.959		.905	
3	.982	.974	.955	.954	.945	.914

that objectivity in scoring is not a problem of any considerable magnitude when dealing with the material provided by the PDPQ.

STUDY I: THE ORGANIZATION OF INTRAPERSONAL LEVELS OF ATTITUDE STRUCTURE AS A FUNCTION OF DIFFERENTIAL OBJECTS OF INQUIRY

This first study was undertaken to determine whether the PDPQ would provide data confirming certain postulated systematic relationships between the personal hypotheses and the expressed reactions in the area of attitude structure (6).

General Hypotheses

If the question-answer process can be formulated in the three-step sequence—(a) the question acts as a stimulus for a personal hypothesis, (b) the hypothesis is checked against the situational context of the stimulus, (c) the verbalized response may represent the hypothesis or be a distortion of the hypothesis to conform to satisfactory adjustment for the individual in terms of situational requirements—then the following systematic relationships between the nature of the question, the personal hypothesis, and the expressed reaction should be demonstrable:

1. Questions dealing with socially conflicted objects of inquiry should have a relatively large discrepancy between personal hypothesis and expressed reaction.

By socially conflicted objects of inquiry is meant objects for which there are normative social expectations of "right" and "wrong" responses, but the expectations have not been satisfactorily internalized by large sections of the population. For example, the Negro is a socially conflicted object of inquiry. Individuals are aware that they are expected to have an attitude

which they may not actually have.

2. Questions dealing with socially neutral objects of inquiry should have a relatively small discrepancy between personal hypothesis and expressed reaction.

By socially neutral objects of inquiry is meant objects for which there are no normative social expectations of "right" and "wrong" responses. For example, the preference for one of two colors is ordinarily a socially neutral object of inquiry.

3. If the personal hypothesis to a socially conflicted object of inquiry is in accord with the situational requirements of the stimulus-situation complex, there should be a relatively small distortion of it in the expressed reaction.

To use the Negro again as an example of a socially conflicted object of inquiry, if the situation of the questioning is a Northern liberal arts college where social expectations are for a positive reaction to the Negro, and the personal hypothesis is positive, the verbalized response will tend to remain positive.

4. If the personal hypothesis to a socially conflicted object of inquiry is not in accord with the situational requirements of the stimulus-situation complex, there should be a relatively large distortion of it in the expressed reaction.

In the foregoing example, if the personal hypothesis is negative, the verbalized response will tend to become positive.

5. If the expressed reaction to a socially conflicted object of inquiry is not in accord with the requirements of the stimulus-situation complex, the personal hypothesis will not be congruent with the requirements of the stimulus-situation complex.

Thus, if the verbalized response to the Negro is negative, the personal hypothesis is negative. If an individual

defies social expectation by admitting to the socially negative reaction, the personal hypothesis is socially negative.

6. If the expressed reaction to a socially conflicted object of inquiry is in accord with the requirements of the stimulus-situation complex, the personal hypothesis may or may not be congruent with the requirements of the stimulus-situation complex.

In the foregoing example, if the verbalized response is positive, the personal hypothesis may or may not be positive. Since there is a social gain in giving the positive socially approved response in this situation, the response may represent a distortion of the internal negative or socially disapproved reaction.

7. A distribution of expressed reactions to a socially conflicted object of inquiry should tend toward a J curve, while a distribution of personal hypotheses to the same object of inquiry should be more symmetrical and approach normality.

The distribution of personal hypotheses conforms to individual differences toward the given object of inquiry; the expressed reactions include distortions of the personal hypotheses to conform to social expectations. The distribution of expressed reactions is representative of the individual differences "squeezed together" in conformity with the norm of expectation (1).

If the PDPQ provides data relevant to the personal hypotheses and the expressed reactions of the model, then the relationships stated here in terms of the question-answer process should hold when restated in terms of differential behavior to the direct and projective instruments. Thus, the relationship stated in hypothesis 4 should be equally applicable when restated as: If the response to a *projective* question regarding a socially

conflicted object of inquiry is not in accord with the situational requirements of the stimulus-situation complex, there should be a maximal distortion of it in the response to the *direct* question. And so on for each of the other hypotheses. Clearly, it is only in terms of observable and mensurable test behaviors that the hypothesized relationships can be verified.

Subjects, Instruments, and Procedures

Subjects

The design of the study required a group of subjects in a community having certain recognizable criteria regarding the major experimental objects of inquiry of the investigation, i.e., attitudes toward parents, the self, religion, and minority groups. The college community, at least as represented by what ought to be given the "Records Office," has such criteria. The situational stimuli emanating, however implicitly, from the office tend to remind the student that there are normative "rights" and "wrongs" with respect to parental attitudes, personal adjustment, and social values. Responses to direct questions regarding these objects of inquiry could be expected to move in predictable directions. For purposes of the study, the college student in the college setting was a good subject, and accordingly a class of 48 freshmen and sophomores in a women's college in New England were utilized as the subjects for the experiment.

Instruments

The projective instrument comprised fifty-nine items of the sort previously described. There were three types of items: (a) socially conflicted objects of inquiry, (b) socially neutral objects of inquiry, and (c) "filler" or decoy items. The content of the items was contingent,

of course, on the hypotheses under test. Thus, in addition to intrinsic interest, the 10 Negro items were chosen as a major focus for analysis because they seemed on a priori grounds to represent an unequivocal example of a socially conflicted area of inquiry. In a New England liberal arts college community there are normative standards as to the proper attitudes to express regarding such things as "Negroes moving in next door" or "having a Negro doctor." The public situation of questioning demands a positive reaction to such inquiries. But there is no lack of evidence for the existence of other less public, but no less real, covert feelings to the contrary regarding these same objects.

Similar considerations underlay the choice of personality, religious, and parental objects of inquiry. The expectations are for self-confidence, religious tolerance, and parental respect, and usually these are forthcoming in the public situation. But again there is no lack of evidence of deep-seated unexpressed feelings to the contrary. It was anticipated that these differences between the overt and covert levels would be reflected in the discrepancies between the direct and the projective responses.

Objects of inquiry for which such discrepancies in specific directions might be anticipated could be chosen easily enough on the basis of the conceptual model. However, if a more rigorous test of the theory and the method was to be made, not only was it necessary to demonstrate systematic discrepancy between the levels of reaction, but it was also desirable to demonstrate predictable *lack of discrepancy*, especially if both discrepancy and lack of discrepancy could be shown to rest on firm theoretical ground.

There are two types of items for which such lack of discrepancy could be antici-

pated. The first represents objects of inquiry which, though controversial, are not socially conflicted. These were the "political" items. There are, to be sure, opinions of "right" and "wrong" regarding political matters, and to this extent these items are not socially neutral as defined in this study. But there is greater latitude for nonconformity regarding these objects than in other areas. Certainly severe sanctions are not as likely to be applied in a college community (at least, not when this study was made) for either a positive or a negative response to a question regarding the desirability of "public ownership" as there might be for a negative response to a question regarding the desirability of religious tolerance.

In addition to this type of controversial but unsanctioned item, a truly socially neutral item was needed. The reasoning behind the construction of such an item involved the following considerations: (a) The item had to elicit some emotional reaction so that the response would not be detached or casual; (b) The item had to consist of a free choice between two emotionally loaded objects so that the scoring might be unequivocal; (c) The objects had to be such that, although emotionally loaded for the respondent, there would be no socially "right" or "wrong" choice—either one would be *right*; the choice is socially neutral; (d) On the technical side, the two objects had to consist of the same number of letters since in a free response test where there is an ostensible premium on speed, the shorter word might be written not because it represented the preferred choice but because it could be written faster.

The item which was constructed and used in the study was: "As between mother's and father's ways Liz usually

chose. . . ." This item seems to meet all of the foregoing requirements. The words "mother's" and "father's" are of equal length, and, although the objects are emotionally loaded, they are socially neutral—either one is a perfectly proper model. In contrast to the socially conflicted objects, the discrepancy between personal hypothesis and expressed reaction for this item should be minimal.

In addition to the experimental items there were decoy or filler items to keep the critical items properly spaced. For example, the Negro items—the items which form the crux of the present study—were presented as items 6, 12, 17, 23, 30, 35, 41, 46, 51, and 56. Questions in the other areas were similarly spaced.

The direct instrument contained forty items matching in content forty items from the projective instrument. All the experimental objects of inquiry were included and such other items as were required for fillers. The wording of the stubs of the direct items was similar to the wording of the matched projective items, except of course that the third person was changed to the first person. In changing the stubs from third to first person, some differences in wording the items occurred in the instruments of the first study; in the second study these differences were eliminated.

Procedures

The projective instrument was administered disguised as a verbal speed test. As stimulation for speed, considerable show was made of announcing the time elapsed at regular intervals and of clocking students as they finished. Sixteen days later the direct instrument was administered as a student census questionnaire. Both instruments were administered during regular class hours by the course instructor.

A copy of the instruments and directions for administration are included in Appendix A.

Results

The results are presented with reference to the general hypotheses formulated above. The hypotheses are restated in substantive terms for the specific objects of inquiry used in the instruments, and the findings are summarized to show whether the anticipated relationships between the projective and direct responses obtained.

Hypotheses 1 and 2

Questions dealing with socially conflicted objects of inquiry should have a relatively large discrepancy between projective and direct responses; the direction of change from the projective to the direct response should be toward the normative expectations for the particular object. Questions dealing with socially neutral objects of inquiry should have a relatively small discrepancy between projective and direct responses. Specifically, there should be large discrepancies for the (a) parental, (b) personal, (c) religious, and (d) racial tolerance objects of inquiry, and negligible discrepancies for the (e) political and (f) the specially constructed socially neutral items.

Parents as objects of inquiry. The change to be expected is from personal hypotheses of dissatisfaction (negative projective responses) to expressed reactions of satisfaction (positive direct responses). The data are presented in Table 2. It will be seen that the hypothesized relationships are clearly confirmed, since the discrepancies are large enough to result in significance beyond the .001 probability level.

Personal competence as object of in-

TABLE 2

FREQUENCIES OF NEGATIVE PROJECTIVE AND DIRECT RESPONSES TO ITEMS CONCERNING *Parents*,
WITH CHI-SQUARE AND *p* VALUES FOR DIFFERENCES^a
(*N*=48)

Object of Inquiry	<i>f</i> Projective Negative Responses	<i>f</i> Direct Negative Responses	χ^2	<i>p</i>
Mother (15, 34)	41	15	23.31	< .001
Father (8, 40)	37	7	27.03	< .001
Total	78	22		

^a All chi squares in this study were computed by means of formula 86a in reference 13. In this and the succeeding tables, the numbers in parentheses following the specification of the object of inquiry refer to the item numbers on the projective and direct instruments. Thus, the item in which "mother" was the object of inquiry was item 15 on the projective instrument and item 34 on the direct instrument.

quity. Included in this area were items which elicited reactions to frustration, reactions to undertaking responsibility, attitudes toward others, feelings of optimism-pessimism, and independence-dependence. The change to be expected is from personal hypotheses of incompetence (negative projective responses) to expressed reactions of competence (positive direct responses). The combined data for eight items in this area are presented in Table 3. Seven of the eight items yield chi-square values that are significant beyond the .01 *p* level. The probability of chance accounting for the difference in response in the expected direction is in-

calculably small, especially since the one-tailed test of significance is appropriate in combination with the predictions of the direction of the change to be expected. The hypothesized relationships are clearly confirmed.

Religious tolerance as object of inquiry. Three items were originally included in this area. One item proved to be inadequately structured and yielded projective and direct responses of only limited comparability. The other two items gave responses of very high comparability. The relevant projective items were phrased as follows: "When the man running for Mayor admitted he did not

TABLE 3

FREQUENCIES OF NEGATIVE PROJECTIVE AND DIRECT RESPONSES TO EIGHT ITEMS CONCERNING
Personal Competence, WITH CHI-SQUARE AND *p* VALUES FOR DIFFERENCES
(*N*=48)

Object of Inquiry	<i>f</i> Projective Negative Responses	<i>f</i> Direct Negative Responses	χ^2	<i>p</i>
Thought job too much (4, 1)	39	8	29.03	< .001
Others are around (16, 9)	27	6	16.00	< .001
Jobs with responsibility (21, 8)	25	8	12.19	< .001
Others doing better (29, 24)	33	8	21.33	< .001
Do whole thing by oneself (39, 31)	30	6	22.04	< .001
Odds are against one (44, 37)	24	9	9.33	< .01
Working with others (52, 28)	13	10	.33	—
Not invited to party (55, 25)	40	15	21.33	< .001
Total	231	70		

TABLE 4
FREQUENCIES OF NEGATIVE PROJECTIVE AND DIRECT RESPONSES TO ITEMS CONCERNING
Religious Tolerance, WITH CHI-SQUARE AND p VALUES FOR DIFFERENCES
($N=48$)

Object of Inquiry	f Projective Negative Responses	f Direct Negative Responses	χ^2	p
Not believe in God (26, 13)	32	9	19.36	<.001
Belong to different religion (57, 26)	17	0	15.06	<.001
Total	49	9		

believe in God, Joan decided . . ." and "When she found out that the candidate belonged to a different religion, Ruth decided. . . ." The change to be expected is from personal hypotheses (projective responses) of intolerance to expressed reactions (direct responses) of tolerance. The data are presented in Table 4 and amply confirm the hypothesized relationship. The difference in response to the two items is noteworthy. Projective responses indicating a negative reaction to a candidate who "did not believe in God" run to about two-thirds of the subjects; to a candidate who "belonged to a different religion" to about one-third. When the same subjects are asked directly, the negative reaction drops to slightly more than one-fifth in the first

case, and to zero in the second case.

The Negro as object of inquiry. A sizable discrepancy between projective and direct responses involving the expression of favorable or unfavorable attitudes toward the Negro might be expected. There were 10 items in this area; the data are presented in Table 5.

Inspection of Table 5 indicates that the difference between personal hypothesis and expressed reaction is statistically significant for each of the 10 items focusing on the Negro—a highly conflicted object of inquiry. These striking data clearly conform to the predicted pattern and serve to confirm the hypothesis.

As in the preceding items, the PDPQ reveals some striking differences in atti-

TABLE 5
FREQUENCIES OF NEGATIVE PROJECTIVE AND DIRECT RESPONSES TO ITEMS CONCERNING THE *Negro*,
WITH CHI-SQUARE AND p VALUES FOR DIFFERENCES
($N=48$)

Object of Inquiry	f Projective Negative Responses	f Direct Negative Responses	χ^2	p
Residing next to Negro (6, 4)	31	14	11.13	<.001
Having Negro forelady (12, 36)	35	9	22.32	<.001
Having a Negro doctor (17, 7)	25	11	8.45	<.01
Dancing with a Negro (23, 10)	28	16	5.04	<.05
Going to school with Negroes (30, 16)	22	11	5.88	<.02
Using Negro beauty shop (35, 33)	28	14	10.56	<.01
Admitting Negroes to club (41, 39)	29	15	8.45	<.01
Hiring many Negroes (46, 10)	27	6	17.39	<.001
Working next to Negroes (51, 27)	21	2	19.05	<.001
Negro in accident (56, 73)	31	1	28.03	<.001
Total	277	99		

TABLE 6

FREQUENCIES OF NEGATIVE PROJECTIVE AND DIRECT RESPONSES TO *Political-Economic* ITEMS,
WITH CHI-SQUARE AND p VALUES FOR DIFFERENCES
($N=48$)

Object of Inquiry	f Projective Negative Responses	f Direct Negative Responses	χ^2	p
Public ownership (53, 22)	20	19	0.00	—
Voting Socialist (24, 11)	26	20	1.13	—
Total	46	39		—

tude structure with respect to the Negro depending on the specific Negro-white interaction that is described. There are, for example, noteworthy variations in the relationship between personal hypothesis and expressed reaction regarding "working next to a Negro" and "residing next to a Negro," between "dancing with a Negro" and "blaming a Negro in an accident." A detailed analysis of these and other substantive findings will be the subject of another report.

Political-economic opinion as object of inquiry. Two items provide data in this area. The expectation was that even with these controversial items there would be lesser change from projective to direct response than with the highly conflicted objects of inquiry. Unpopular political opinions, especially in the college community, are less likely to constitute a personal or social threat than unpopular opinions about religious tolerance or adverse attitudes toward one's family. There is, however, one technical difficulty in this area—or in the wording of the items—in that the questions elicited a great many "don't know" or otherwise unclassifiable answers. Approximately 30% of the responses were of this nature. Nevertheless, discounting these indeterminate responses, there are sufficient data to permit ready interpretation.

The two items in this area were stated:

"When she was asked her opinion of public ownership, Kate said . . ." and "When Connie was asked whether she could ever see herself voting Socialist, she said. . . ." Table 6 indicates that there is virtually no change from the projective to the direct responses elicited by the items, and thereby sustains the hypothesized relationship.

Socially neutral choice as object of inquiry. Projective item 34 reads: "As between mother's and father's ways Liz usually chose. . . ." The alternative offered is emotionally loaded, but, whatever the choice, there is no contrary social sanction. The hypothetical expectation was that the discrepancy between the projective and the direct response would be negligible.

The gathering of the data for this section required one variation from the usual procedure. When the direct instrument was administered, many of the subjects responded with "both," "neither," or some such evasion. Only 18 of the 48 subjects responded in such a way as to permit study of the congruence or disparity of the direct and projective responses. The data for these 18 cases confirmed the hypothesis of no change. However, the meagerness of this sample indicated the desirability of additional data. Hence, the open-ended direct item was changed to a multiple-choice form where

the subjects could not give a neutral response, but were forced to give preference to either mother or father. The direct multiple-choice item was: "As between mother's and father's way, I choose (a) practically always mother's, (b) usually mother's, although sometimes father's, (c) practically always father's, (d) usually father's, although sometimes mother's." In addition, it was also decided to include a sample of 32 male college students. The projective item for males read: "As between mother's and father's ways, Carl usually chose. . . ." The direct item was the same for both males and females. Table 7 presents the relevant data.

The chi-square values reported in Table 7 are clearly not significant, and the data may thus be interpreted as indicating that when the object of inquiry is socially neutral there is no significant difference between the personal hypothesis and the expressed reaction. The hypothesis that the discrepancy between projective and direct responses should be relatively small is clearly confirmed.

Hypotheses 3, 4, 5, and 6

The preceding section explored the general relationship between projective and direct responses as the content of

the question is changed. It was shown that the relationship is systematic: the discrepancy between responses will be large when the content of the question is socially conflicted and small when socially neutral.

The present section will explore the behavior of projective and direct responses to a socially conflicted object of inquiry—the Negro—focusing on the dynamics of the discrepancy between the levels revealed by the two instruments.

The hypothesized relationships may be represented in the following fourfold table, where a plus (+) stands for a response in accordance with social expectations (a positive or at least a non-negative response to a Negro item) and a minus (−) for a response not in accord with social expectations (a negative response hostile to the Negro). The prediction from the hypothesized relationships is that the lower left-hand cell (C) should have the smallest number of frequencies—ideally, none. That is, no projective positive responses should be distorted to direct negative responses, no direct negative responses should be distortions of projective positive responses. In other terms, if the personal hypothesis is unprejudiced there is no reason for a respondent—under the conditions of this

TABLE 7

COMPARISON OF PROJECTIVE AND DIRECT RESPONSES TO ITEM CONCERNING PREFERENCE FOR MOTHER'S OR FATHER'S WAYS, FOR MALE, FEMALE, AND TOTAL SAMPLES, WITH CHI-SQUARE VALUE FOR THE CHANGE BETWEEN RESPONSES

		Females Projective Father Mother		Males Projective Father Mother		Total Projective Father Mother			
Direct	F	7	4	F	18	1	F	25	5
	M	3	26	M	4	9	M	7	35
		$\chi^2=0.00$		$\chi^2=0.80$		$\chi^2=0.08$			

investigation—to change to a prejudiced expressed reaction.

		Personal Hypothesis (Projective Response)	
		+	-
Expressed Reaction (Direct Response)	+	A X	B X
	-	C O	D X

Actually, of course, there were many more responses in cells A and B than in C and D. In all, negative direct responses to the 10 items totaled 99, about 20%, while negative projective responses totaled 277, or about 58%. But if the postulated relationships hold, then there should be fewer responses in cell C than in any of the others, including cell D. Table 8 presents the data for this phase of the study.

These data confirm the predicted relationships. Projective positive responses only rarely (about 10% of the cases) were distorted to direct negative responses. Projective negative responses were frequently (72% of the cases) distorted to direct positive responses. A direct positive response may or may not coincide with the subject's projective response—

TABLE 8
RELATIONSHIP BETWEEN NEGATIVE AND POSITIVE DIRECT AND PROJECTIVE RESPONSES TO 10 NEGRO ITEMS (N=48)

		Projective			
		+	-		
Direct	+	182 A	199 B	381	
	-	21 C	78 D		
		203	277	480	

there were 48% such agreements. A direct negative response frequently coincided with the subject's projective response—there were 78% such agreements. This analysis of the data supports the hypotheses and indicates that the direct and projective responses behave quite like the personal hypotheses and expressed reactions would be expected to change, and as one would expect public and private attitudes to vary.

Hypothesis 7

The final hypothesis to be examined in this phase of the investigation postulated that the distribution of personal hypotheses should reflect individual differences in private attitudes toward the given object of inquiry, while the distribution of expressed reactions should represent a distortion in terms of closer conformity to social expectations. Hence, it was expected that the distribution of negative projective scores to a socially conflicted object of inquiry should approximate normality, but that the distribution of negative direct scores should tend toward a J curve.

For this section of the study subjects were assigned scores in terms of the number of negative responses to the 10 projective items and the corresponding 10 direct items which deal with the Negro. The graphic representation of the projective and direct distributions is presented in Fig. 1.

Inspection of the modal values and spread of scores in these two distributions suggests the appropriateness of the hypotheses advanced. The modal score of the projective test distribution is 7, and the frequencies are distributed in generally decreasing order toward the extreme score values of 0 and 10. In contrast, the modal score of the direct test distribution is 0, and the frequencies

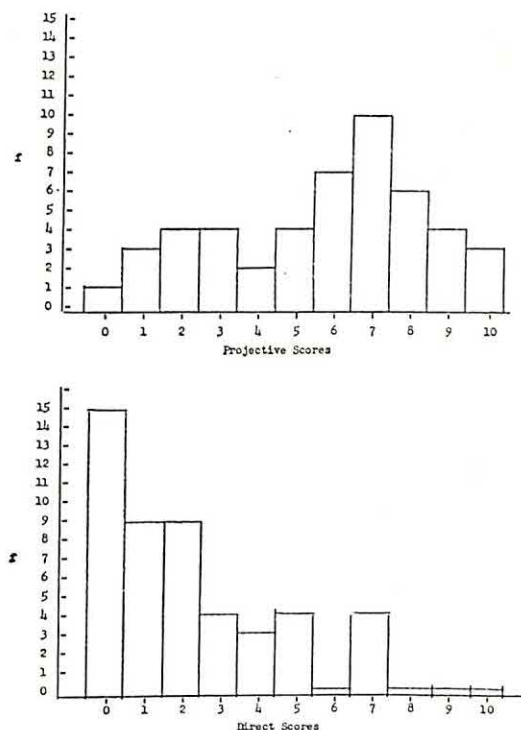


FIG. 1. Distribution of Projective and Direct Scores on 10 Negro Items.
($N = 48$)

diminish as the score values increase. This is the general pattern of the J curve and the data are therefore interpreted as confirming the hypothesized relationships.

STUDY II: THE ORGANIZATION OF INTRA-PERSONAL LEVELS OF ATTITUDE STRUCTURE AS A FUNCTION OF DIFFERENTIAL PATTERNS OF SOCIALIZATION

The first study provides clear evidence that the method of Paired Direct and Projective Questionnaires can profitably be applied to the systematic investigation of overt and covert levels of attitude structure. Can the method also be applied profitably to the investigation of variations in the development of these levels of attitude as a function of such factors as age, sex, sibling position, and social class status? This is the problem

under consideration in the second study, and it is explored within the general framework of the concept of socialization (16).

Socialization is conceived here as a twofold process. On the one hand, the child learns to renounce, suppress, or re-direct drives and behavioral impulses that are at variance with proper social standards. On the other, he learns to bring his observed behavior into conformity with the norms and expectations of the society in which he lives. It is a process by which the "persona" is differentiated from the "self," and the overt response from the covert impulse. The infant inhibits impulse but little; the adult a great deal. This is not to say that socialization involves the complete elimination of unacceptable behavioral tendencies. The socialized child has not so much eradicated forbidden inclinations as learned to restrain their manifest expression. He has learned, in Allport's terms, that "in everyday life it is often necessary to mask the true self, to present to the world the appearance of conventional fitness" (2, p. 39). It is during the period of socialization that the distinctive individual patterns of articulation between these covert and overt levels of attitude are established. Indeed, from this point of view, socialization may be defined as the successive differentiation and organization of intrapersonal levels of reaction.

So considered, the process of socialization must be studied in terms of three major concepts: (a) covert behavior, or the *personal hypothesis* of the levels model; (b) overt behavior, or the *expressed reaction* of the model; and (c) the congruence or discrepancy between the two, or as it is referred to in the context of this phase of the study, the *Index of Differentiation*.

The PDPQ produces the three measures needed to study socialization within this framework. With respect to any object of inquiry it provides: First, a measure of the covert attitude, contained in the personal hypothesis elicited by the projective instrument; second, a measure of the overt attitude, contained in the expressed reaction elicited by the direct instrument; and finally, a measure of the discrepancy between the two, and this is the third and crucial metric, the Index of Differentiation (ID).

If the conceptual framework is tenable and if the method does in fact provide the necessary measures, then the following major relationships between the covert and overt levels of attitude structure as reflected by systematic variations in the ID for different age, sex, sibling position, and social class groups should be observed.

General Hypotheses

1. The mean Index of Differentiation should increase as a direct function of age. Socialization is a process in time, and individuals—in this case, children of school age—subjected to socializing pressures for longer periods should have greater ID scores than those under such pressure for lesser periods.

2. The mean Index of Differentiation should be greater for girls than for boys in the same age group. The greater social conformity and maturity of girls over boys during the years adjacent to the pubertal period is a well documented finding of developmental psychology, and this should be reflected in the ID metric.

3. The mean Index of Differentiation should be greater for the only child in a family than for the oldest of three children, which should in turn be greater than for the youngest of three children in a family. There is evidence from de-

velopmental studies for this relationship (11), and if the concepts and methods are sound, then the indicated relationship should be reflected in the respective ID scores.

4. Finally, the mean Index of Differentiation should be significantly different for children of the higher or middle socioeconomic classes than for children of the lower class. Numerous studies show the varying pressures for inhibition of impulse and for conformity to distinct standards required of children in the different socioeconomic classes. Again, if the conceptual structure and method are sound, these differences should be reflected in the ID metric.²

Subjects, Instruments, Procedures, and Methods of Analysis

Subjects

Eight schools participated in the present study. Two of these schools are public elementary schools in a lower-middle-class industrial suburb of Chicago. The third school is a private, nonsectarian school for boys located on the south side of the city of Chicago. Typically, the fathers of boys attending this school are professional or business men. The fourth

² The key used in scoring the instruments of this study undoubtedly reflects middle-class norms. Theoretically, this constitutes a "bias" in favor of the subjects representing the same social class. However, the test of the technique requires that all subjects be judged by the same standard. This is so because the critical question in this phase of the study asks whether or not the technique is adequate for the detection of *differences* in the behavioral patterns of different social classes. Obviously these differences can manifest themselves only if the same criterion is applied to all subjects. Different scoring keys for different social classes could provide valuable information relating to differences *within* these classes, but this should hardly be attempted until the technique has proven its worth in terms of the differences *between* social classes. It is on this latter problem that the present investigation focuses.

school is a private, nonsectarian school for girls, serving the same social class as the private school for boys. Three of the schools included in the sample are Catholic parochial schools, all located on Chicago's south side. One of these schools is situated in what is considered to be one of the most desirable residential areas, and most of the children attending this school are from families in the professional class. The second parochial school is located in a transition area, and until two years prior to the testing for this study was attended typically by white children from lower-middle-class families. At the time of testing this school was composed of approximately equal numbers of white and Negro children. The third parochial school is located in a slum and federal housing project area, and approximately 95% of the children attending this school are Negro. The final school included in the sample is a Lutheran parochial school located in a middle-class suburban residential community west of Chicago.

Since the experimental method presupposes a knowledge of reading and writing, it was deemed unsuited to children below the third grade level. Hence, the third grade was chosen as the lower limit and the sampling was made in such a way as to have the greatest representation over the age range 8 to 13. Table 9 presents a breakdown of the sample according to age and sex.

Negro children constitute 11% of the sample, and the sample is divided according to religious preference as follows: Catholic, 42%; Jewish, 6%; Protestant, 47%; and 5% of the sample admit no religious affiliation.

Instruments

Three areas of research interest were chosen for inclusion in the instruments

TABLE 9
DESCRIPTION OF SAMPLE

Age	Boy	Girl	Total
8	70	63	133
9	78	88	166
10	76	98	174
11	66	72	138
12	64	73	137
13	74	91	165
Total	428	485	913

employed in this phase of the study. The first area may be described as the *family* area, and the items in this area pose situations designed to elicit attitudes toward parents and siblings. The second area is called the *personal* area, and these items were constructed to obtain reactions to situations dealing with self-confidence, frustration, relationships with members of the opposite sex, and general value orientations. The *social* area was the third area included in these instruments, and the items here relate to such things as attitudes toward racial and religious groups.

In all, four instruments were employed in this phase of the study. These tests are designated as Form B, Form G, Form IB, and Form IG. Forms B and G are the projective tests for boys and girls respectively. These forms differ in the names that are used in the third person references, but the substance of the items in the tests is identical. Thus, in Form B, item 9 reads: "Don sometimes thinks his family. . . ." In Form G the item is phrased: "Dot sometimes thinks her family. . . ." Forms IB and IG are the direct forms of the test, in which the items are phrased in the first person. The use of two direct instruments was necessitated by the inclusion of item situations referring to members of the opposite sex. Thus, for example, item 11 in Form IB is phrased: "If I had to work with girls

all the time, I" In Form IG the item takes the form: "If I had to work with boys all the time, I"

Each of the four forms of the test consists of 40 incomplete sentences, with the items being presented in the same numerical order in all of the instruments. However, only items 3 through 40 are actually scored in the present study; items 1 and 2 were designed to serve as innocuous practice items although they were not explicitly labeled as such. Copies of the instruments and directions for administration are included in Appendix B.

Procedures

The projective instrument was administered first, and the direct test was administered after a two-week interval. At the end of the second test session the children completed a personal data sheet which provided data concerning age, sibling position, and other variables needed in the analysis of the data.

Analyses

The status variable chosen for use in the study was the socioeconomic factor of occupation of parent, and consequently the analyses to be reported should not be interpreted as standing in perfect relationship to the more rigorous conception of social class. Direct correspondence of occupational groups and the various social classes simply does not obtain, and hence in this study comparisons will be between socioeconomic groups (as identified by occupational groupings), rather than between social classes as otherwise defined. Data provided by Warner (17, p. 165) indicate that of a total of 110 people classified in the top three occupational groups, 109 were in the middle or upper class. On the other hand, 69 out

of 98 people classified in occupational groups 4 through 7 were in the lower class. These data suggest that pooling the first three occupational groups establishes the bounds for the upper and middle classes, while a combination of the last four occupational classes roughly delimits the lower social class. This dichotomy has been utilized in the analyses to be presented, and although references in the text will be to the high and low socioeconomic groups, the reader may—within the limits of error indicated above—equate these to the middle and lower social classes.

The instruments were scored in terms of a positive-negative dichotomy, as in the preceding study. Each subject was assigned three scores. The projective score and the direct score represent the number of negative completions on those instruments. The third metric—the Index of Differentiation—represents a proportion. This Index has been defined as the proportion of negative responses on the projective instrument which is reversed to positive responses on the direct form of the test. In a pilot study a comparison was made of the advantages of computing this proportion in terms of item-by-item reversal, or merely in terms of the difference between the projective score and the direct score. The correlation between the Indexes computed by these two different methods was found to be .97. The method based on the difference between the two scores is considerably more economical of time than the item-by-item method, and hence was adopted for use in the present study. The computational procedure for this Index is expressed by the formula

$$ID = \frac{P - D}{P}$$

where P represents the score on the

projective test, and D the score on the direct test.

The variables of sex, age, birth rank, and socioeconomic status were the dimensions used in analyzing the three sets of scores. The problem of controlling variables becomes crucial when it is desired to assess the influence of each of these variables with the others held constant. Under the conditions of sampling which characterized this study, complete experimental control of the variables was impossible. Hence, statistical control, as exemplified by the covariance technique, was used. However, when dealing with discrete variables such as sex it was also necessary to achieve control in terms of the subjects selected for inclusion in some of the subsamples on which the analyses were based. Consequently, some of the analyses to be reported are based upon the data derived from the total sample, while others are based upon randomly selected subsamples of the total group.

Results

The findings of this study are presented in terms of the four major hypotheses advanced above. These hypotheses are oriented around the Index of Differentiation, since it is the critical metric in this phase of the investigation. In a sense, the Index represents an attempt to express the extent of socialization in objective terms by indicating the proportion of asocial covert behavioral tendencies which are censored and denied expression at the overt level. It summarizes the relationships existing between the personal hypotheses and expressed reactions, and its value transcends the worth of the separate measures which—when used conjointly—serve as the means for its calculation. Hypotheses relating to the Index of Differentiation constitute the most rigorous test of the

TABLE 10
MEAN ID SCORES FOR AGE GROUPS

Age	N	Mean ID	
8	133	29.02	29.00
9	166	28.98	
10	174	36.91	37.87
11	138	39.08	
12	137	41.01	40.87
13	165	40.76	

PDPQ, since verification of these hypotheses involves, at the same time, the fulfillment of certain inflexible relationships between the direct and projective responses. Accordingly, in interpreting the data of the study, emphasis will be placed on the hypothesized patterns of the ID scores. However, data for the other two metrics will also be presented, chiefly for the purpose of indicating the nature and direction of the discrepancies summarized in the ID scores.

Hypothesis 1

The mean Index of Differentiation should increase as a direct function of age.

Table 10 presents the mean ID scores for the different age levels, in both one-year and two-year groupings.

It is readily apparent that the over-all increase from age 8 to age 13 is of considerable magnitude and in the predicted direction. The presentation of the data in the form of one-year age intervals is adopted in the interest of convenience and completeness. It is not necessitated by the hypothesis under test, since the hypothesis referred to the totality of the distribution and the over-all trend, rather than to the discrete arbitrary segments represented by one-year age classes. When the data are grouped in two-year

TABLE 11
MEAN PROJECTIVE AND DIRECT SCORES
FOR AGE GROUPS

Age	N	Projective	Direct
8	133	18.04	13.23
9	166	17.33	12.40
10	174	18.62	11.71
11	138	18.73	11.12
12	137	18.83	10.99
13	165	18.73	10.83

intervals, the progression is striking and unmistakably in conformity with the hypothesis, and the analysis of variance yields an F of 8.14 (p beyond the .01 level).³

The ID metric summarizes the relationship between the personal hypotheses and expressed reactions for children of different ages, but it does not reveal the dynamics of that relationship. It cannot be determined, for example, whether the larger ID scores for older children are produced by an increase in projective scores coupled with stability in direct scores, or by an increase in direct scores accompanied by no change in projective

³ Since ID scores are expressed as proportions, it was necessary to utilize a transformation in order to make these scores amenable to valid analysis by the F -ratio technique. Consequently, the ID scores were processed by means of the arc-sine transformation, and the analyses of ID scores have been made on the basis of these transformed values (10, p. 594). The L_1 test of homogeneity of variance was made for all analyses of variance reported in this study.

With respect to the age hypothesis under test here, we ought perhaps to add that, at this point, in order to examine any possible confounding of the age and sex variables which might have been introduced by the slightly unequal division of boys and girls in the several age intervals of the original sample, a second analysis of the ID scores was made. In this second analysis, 421 boys and 421 girls were randomly selected by age groups from the parent sample and divided equally by sex at each age level. The analysis of ID scores for this sample resulted in an F of 7.92 (p beyond the .01 level), thus confirming the original analysis.

scores, or by decrease in one variable combined with increase in the other. The nature of the relationship can be understood by referring to the distributions of the direct and projective scores which are the basis of the Index of Differentiation. Table 11 presents the necessary data.

It is clear that the projective scores manifest a high degree of stability—the difference in mean projective scores for 8-year-old children and 13-year-old children is only .69. The direct scores, on the other hand, reflect a pattern of gradual but consistent decrease from one age to the next higher age. The increase in ID as a function of age results, therefore, from the covert reactions remaining relatively stable while the overt reactions move systematically in the direction of greater social conformity.

This finding appears to warrant the interpretation that the so-called “private” levels of reaction are not subject to any real modification over the given age range as a result of the learning experiences subsumed under the concept of socialization, at least not with respect to the objects of inquiry represented in the present instrument. The learning that goes on is reflected instead in changes at the “public” levels of reaction. The child learns to inhibit existing socially disapproved personal hypotheses, but these do not thereby cease to exist. They are only overlaid with socially approved expressed reactions.

Hypothesis 2

The mean Index of Differentiation should be greater for girls than for boys.

The relevant data are summarized in Table 12. It will be noted that the data here are based on 884 cases instead of the 1,033 cases used for the age group variable. In analyzing the sex group data it

TABLE 12

MEAN INDEX OF DIFFERENTIATION, PROJECTIVE, AND DIRECT SCORES FOR SEX GROUPS

Metric	Sex Group		F	p
	Boys	Girls		
ID	32.18	39.43	8.23	<.01
Projective	19.15	17.67	16.44	<.01
Direct	12.63	10.85	25.00	<.01
N	428	456		

was desirable to include a test of the significance of the interaction between the sex and the age variables. Such a test required that the analysis of variance be made on the basis of a twofold classification. Unequal numbers of cases in each of the categories poses no problem in a onefold classification as employed with the age data, but does represent complications for higher order classifications. The difficulty can be overcome by taking equal numbers of random samples for each group, but in such procedure a considerable amount of the available data is usually eliminated. To obviate this problem it was decided to utilize an approximation method of analysis of variance for unequal frequencies in the subclasses of two classifications as presented by Johnson (9, p. 261). Although this procedure involved the elimination of some of the cases, the loss was considerably less than would have resulted by taking equal random samples for each group.

The twofold analysis of variance design provided a test of the significance of the means and a measure of the interaction between age and sex for each of the three metrics. The three interactions proved insignificant. The difference between the ID scores of the boys and girls was significant, and in the predicted direction. The second hypothesis is substantiated by the data of the study. Clearly, the discrepancy between the per-

sonal hypotheses and expressed reactions is greater for girls than for boys between the ages of 8 and 13.

It is important to note that, although the mean ID score is greater for girls than for boys, the reverse relationship is true for the direct and the projective scores. That is, although the asocial private and the asocial public attitudes are greater for boys than for girls, the *proportional change* from the socially unapproved personal hypothesis to the socially approved expressed reaction is greater for girls than for boys.

Synthesizing these findings with those from the age relationships, it appears that while there are substantially more asocial behavioral tendencies in the private worlds of boys than of girls, within the same sex group the number of these tendencies does not undergo change but remains surprisingly stable over the years 8 through 13. This implies that, by the time the child has reached the age of 8, he has already acquired a set of personal hypotheses which are typed on the basis of sex and which are not subject to alteration during the next five years. The expressed reactions do undergo consistent modification in the direction of greater social acceptability as a function of age. However, the extent of modification varies for boys and girls, with girls significantly more compliant to social expectations. As has already been pointed out, the proportion of change from socially unacceptable covert responses to socially acceptable overt responses is greater for girls than for boys, and to this extent, of course, girls may be said to be "more socialized" than boys of the same age group.

Hypothesis 3

The mean Index of Differentiation should be greater for the only child than

B read of Eval & Psyl. Research

9-3-81

TABLE 13
MEAN INDEX OF DIFFERENTIATION, PROJECTIVE, AND DIRECT SCORES FOR BIRTH RANK GROUPS

Metric	Birth Rank Groups			<i>F</i>	<i>p</i>
	1/1	1/3	3/3		
ID	40.61	37.21	36.87	3.06	<.05
Projective	19.40	18.80	18.89	1.09	>.05
Direct	10.89	11.58	11.79	1.72	>.05
<i>N</i>	88	88	55		

for the oldest of three children, which should in turn be greater than for the youngest of three children.

The data yielded by the technique are again in agreement with the hypothesized relationships. Table 13 summarizes these data. The Index of Differentiation shows systematic and significant variation for the sibling positions as expected from socialization theory and relevant developmental studies.

The data of Table 13 provide an important illustration of the merit of the technique under test. It must be noted that if only the projective instrument had been employed, the conclusion would be that no significant relationship existed for the sibling position variable. Similarly, if only the direct instrument had been employed, the null hypothesis would have been accepted for this variable. But when both instruments are used conjointly and a metric developed to express the relationship existing between the two scores, this third metric—which is the core of the technique under study—does indicate the existence of a significant relationship.

Hypothesis 4

The mean Index of Differentiation should be significantly different for children of the higher or middle socioeconomic classes than for children of the lower classes.

The data presented for Hypotheses A and B indicated the existence of significant ID differences as functions of age and sex. In seeking to determine the influence of the socioeconomic class variable on the metrics of the study, it became necessary to control the age and sex factors. The age factor, representing a continuous variable, may be controlled statistically by the analysis of covariance technique, and this procedure was adopted. However, sex is a discrete variable and does not lend itself to control by the covariance method. Control of the sex factor was achieved experimentally by using separate samples of boys and girls and doing separate covariance analyses for each sample. These analyses are based on the scores of randomly selected subsamples of 30 cases each for boys and girls for each of the two socioeconomic groups. In order to achieve greater precision for the test of the hypothesis relating to socioeconomic class, the factor of intelligence was held constant by the covariance technique. The analyses are summarized in Table 14.

The data indicate that for both boys and girls significant differences in mean ID scores are obtained for subjects of different socioeconomic classes. The hypothesis is confirmed. The extent to which boys and girls, with age and IQ constant, tend to inhibit their socially unacceptable personal hypotheses is

TABLE 14
MEAN INDEX OF DIFFERENTIATION, PROJECTIVE, AND DIRECT SCORES FOR BOYS AND
GIRLS OF TWO SOCIOECONOMIC CLASSES

Sex	Metric	Socioeconomic Class		<i>F</i>	<i>p</i>
		High	Low		
Boys	ID	49.80	31.60	5.15	<.05
<i>N</i> High = 30	Projective	18.33	18.13	3.82	>.05
<i>N</i> Low = 30	Direct	9.17	12.70	4.48	<.05
Girls	ID	37.97	44.20	4.04	<.05
<i>N</i> High = 30	Projective	16.53	18.00	1.52	>.05
<i>N</i> Low = 30	Direct	10.27	10.10	—	—

clearly a function of the socioeconomic class membership of the children.

It is worthy of special note that the pattern of socialization is different not only for the socioeconomic classes under study, but differs significantly for boys and for girls within the two socioeconomic groups. The mean ID score for boys of the high status group is over 18 points greater than the mean ID score for boys in the low group. This large discrepancy in ID score corresponds to the frequent observation that lower status boys consider the higher status boys "sissies," and the latter in turn consider the former "toughs." The pattern for girls is the reverse. Girls in the low status group change 44% of their asocial personal hypotheses, while the high status girls change 38%. That is, while for boys the ID score of the "highs" is greater than that of the "lows," for girls the ID score for "lows" is greater than that of the "highs," although the difference is not by any means as large for the girls as for the boys. The finding with respect to girls is again consistent with previously reported observations and theoretical considerations. Girls of all classes at the age group under study tend to approximate adult standards of behavior more closely than do boys, and this is especially so at the lower status levels.

SUMMARY AND CONCLUSIONS

There are several lines of cleavage in the evolution of current psychological techniques. In the conceptual realm, the cleavage has been between mechanistic and dynamic models of human behavior; in the substantive realm, between recording observed habits and eliciting underlying motives; in the methodological realm, between asking direct questions and posing projective stimuli. These cleavages are reflected in the diversity of direct and projective instruments brought to bear on both clinical and research problems—instruments that frequently provide a variety of data without any conceptual, substantive, or methodological common denominator.

MacKinnon has summarized the situation as follows:

And so it is that we have had traditionally two kinds of diagnostic tests—on the one hand, tests that reveal (of course, with a large probable error) what a person has done, how he has felt, or now feels, what he has thought or now thinks, in short, the person as a phenotype; and on the other hand, tests that reveal what we have been wont to call the underlying dynamics of the person, in other words, the person as a genotype. But between the two there has been a tremendous gulf to be bridged. . . . We have not yet discovered the formula which will permit us to move from one to the other (12, p. 7).

The method of Paired Direct and Pro-

jective Questionnaires (PDPQ) is proposed as a fruitful technique for bridging the indicated gulf. Analysis suggests that the "formula" must include two essential elements. Theoretically, it must be based on a conceptual framework capable of relating in a psychologically meaningful way the data furnished by the two classes of instruments. Formulation of the question-answer process in terms of *levels* of personality provides such a framework by permitting the response to a direct question to be viewed as emanating from the overt level of behavior, and the response to a projective question as emanating from the covert level of behavior. One is the *expressed reaction*, the other the *personal hypothesis* of the question-answer process, the two levels of response being related in systematic and empirically testable ways. Methodologically, the formula must include a technique that yields direct and projective responses with respect to the *same* objects of inquiry in *comparable* terms so that the theoretically postulated relationships between expressed reactions and personal hypotheses can be plotted.

The PDPQ meets both these requirements. Theoretically, it is based on a formulation of the question-answer process in terms of levels of intrapersonal behavior, and certain invariant relationships between the levels are derived from the formulation. Methodologically, it makes use of directly and projectively stated sentence-completion-type questions in such way that the responses provide direct and projective data to the same objects of inquiry in readily comparable terms.

Two complementary studies were designed to test the applicability of the PDPQ, one dealing with the relationship between intrapersonal levels of attitude structure as a function of *variations in*

the objects of inquiry, the other with the relationship between intrapersonal levels of attitude structure as a function of *variations in socialization* represented by the age, sex, sibling position, and social class variables. In each study certain relationships between the personal hypotheses and the expressed reactions were postulated on the basis of current attitude and socialization theory, and the proof of the PDPQ was predicated on whether the obtained direct and projective responses would conform to these relationships.

Specifically, the first study applied the PDPQ to testing seven postulated relationships, falling into the following three general areas:

1. Questions dealing with socially conflicted objects of inquiry should have relatively large discrepancies between the personal hypothesis (projective response) and the expressed reaction (direct response); the direction of change from the personal hypothesis to the expressed reaction should be toward the normative standard of responses for the particular object. Questions dealing with socially neutral objects of inquiry should have relatively small discrepancies between the personal hypotheses and the expressed reactions.

2. If the personal hypothesis to a socially conflicted object of inquiry is in accord with the requirements of the situation, there should be a relatively small distortion of it in the expressed reaction. If the personal hypothesis to the same object of inquiry is not in accord with the requirements of the situation, there should be a relatively large distortion of it in the expressed reaction. Conversely, if the expressed reaction to a socially conflicted object of inquiry is not in accord with the requirements of the situation, the personal hypothesis will not be in

accord with the requirements. If the expressed reaction to the same object of inquiry is in accord with the requirements of the situation, the personal hypothesis may or may not be in accord with the requirements.

3. A distribution of expressed reactions to a socially conflicted object of inquiry should tend toward a J curve, while a distribution of personal hypotheses to the same object of inquiry should be more symmetrical and approach normality.

The second study employed the two major concepts of the first study, i.e., the personal hypothesis as represented by the projective response and the expressed reaction as represented by the direct response, and developed a third metric, the *Index of Differentiation*. This Index is a measure of the magnitude of the discrepancy between the personal hypothesis and the expressed reaction, and in a sense represents the extent of socialization. The following four hypotheses with respect to variations in the Index of Differentiation as a function of the age, sex, sibling position, and social class status for children 8 to 13 years old were tested:

1. The mean Index of Differentiation should increase as a direct function of age.

2. The mean Index of Differentiation should be greater for girls than for boys in the same age group.

3. The mean Index of Differentiation

should be greater for the only child than for children with siblings.

4. The mean Index of Differentiation for children of the higher or middle socioeconomic classes should be significantly different from that of the lower socioeconomic classes.

The hypothesized relationships of both studies were confirmed without exception, and in the larger proportion of cases at very high levels of confidence. This congruence between the empirical data provided by the PDPQ and the predicted relationships derived from attitude and socialization theory attests to the usefulness of the technique.

Although the PDPQ is surely not the ultimate formula for moving from one level of intrapersonal reaction to another as represented empirically by responses to direct and projective instruments, it does bridge part of the gulf, and thus offers the means of exploring behavioral relationships largely neglected for want of an appropriate method. If the first stage in instrument-making centered on the direct instrument, and the second on the projective one, the PDPQ may very well be considered a third stage by centering on an integration of the first two. In this sense, the PDPQ may be looked upon as a further developmental step in the evolution of objective techniques for the systematic study of human behavior, or at least as pointing to a promising direction for such step.

REFERENCES

1. ALLPORT, F. H. The J-curve hypothesis of conforming behavior. *J. soc. Psychol.*, 1934, 5, 141-183.
2. ALLPORT, G. W. *Personality: A psychological interpretation*. New York: Holt, 1937.
3. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake and G. V. Ramsey (Eds.), *Perception: An approach to personality*. New York: Ronald, 1951.
4. CAMPBELL, D. T. The indirect assessment of attitudes. *Psychol. Bull.*, 1950, 47, 15-38.
5. FREUD, S. *A general introduction to psychoanalysis*. New York: Garden City, 1943.
6. GETZELS, J. W. The assessment of personality and prejudice by the method of paired di-

- rect and projective questionnaires. Unpublished doctoral dissertation, Harvard Univer., 1951.
7. GETZELS, J. W. The question-answer process: a conceptualization and some derived hypotheses for empirical examination. *Publ. Opin. Quart.*, 1954, **18**, 80-91.
 8. HANFMANN, E., & GETZELS, J. W. Studies of the sentence completion test: I. *J. proj. Tech.*, 1953, **17**, 280-294.
 9. JOHNSON, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.
 10. KELLEY, T. L. *Fundamentals of statistics*. Cambridge: Harvard Univer. Press, 1947.
 11. MACFARLANE, J. W. Study of personality development. In R. G. Barker, J. S. Kounin, and H. F. Wright (Eds.), *Child behavior and development*. New York: McGraw-Hill, 1943.
 12. MACKINNON, D. W. Psychodiagnosis in clinical practice and personality theory. *J. abnorm. soc. Psychol.*, 1949, **44**, 7-13.
 13. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
 14. ROSENZWEIG, S. Levels of behavior in psychodiagnosis with special reference to the picture frustration study. *Amer. J. Orthopsychiat.*, 1938, **8**, 622-626.
 15. SANFORD, R. N., ET AL. Physique, personality, and scholarship: a cooperative study of school children. *Monogr. Soc. Res. Child Developm.*, 1943, **8**, No. 1.
 16. WALSH, J. J. The technique of paired projective and direct sentence completion tests applied to the study of socialization. Unpublished doctoral dissertation, Univer. of Chicago, 1953.
 17. WARNER, W. L. *Social class in America*. Chicago: Science Research Ass., 1949.

(Accepted for publication October 8, 1956)

APPENDIX A

INSTRUMENTS USED IN STUDY I

THE MERRILL-BROWN VERBAL INTELLIGENCE TEST

Form 11F

Name:

Age:

Education:

Occupation:

Date:

This is a test of how fast you can think. Complete each of the following sentences so that it makes the best sense possible. You may use either a word or a phrase, although a phrase is preferable. Since you will have only a limited time, work quickly. In most cases the best way to answer the test is to put down the first thing that comes to your mind.

Example:

The unfinished sentence: "When Ann scored the goal, she . . ." may be finished by adding "felt good" or "won the game" or "told her friend" or "wanted to score another" or something like that.

The score you get depends on how quickly you finish the sentences. Work fast. *Do not skip any sentences.*

This is a speed test. Work fast. Do not skip any sentences.

1. The thing Mary liked to do best in her spare time was
2. After Florence finished talking she thought that she
3. Barbara thinks most politicians are
4. When Joan thought the job was too much for her, she
5. What Dot disliked most about her family was
6. When Alice saw Negroes moving in next door, she
7. When they asked Elizabeth to be in charge, she
8. Ruth sometimes thought her father was
9. When Marge heard she would have to join a union in her new job, she
10. Kate often thinks of herself as a
11. When Anne's new friend said she was against God, Anne
12. When they put a Negro forelady over her, Edna
13. The reason Lillian tried to get ahead was
14. Nancy believes most people think of her as
15. When they tried to get Margaret to be like her mother, she
16. When Janet had something to say and others were around, she
17. When Jane saw the doctor they were trying to give her was a Negro, she
18. According to Irene, if a person remains poor it is because
19. When they asked Louise what nations the U. S. could trust, she said
20. Like most girls, Betty sometimes thinks her family is
21. Most jobs with responsibility made Nora feel
22. From past experience, Gail felt most people were
23. When Jean saw her younger sister dancing with a Negro, she
24. When Connie was asked whether she could ever see herself voting Socialist, she said
25. Above all else, Mary wished her father were

26. When the man running for mayor admitted he did not believe in God, Joan decided
27. After Madeleine finished the test, she thought that she
28. If Ruth's mother could change things she would try to
29. When Alice saw others doing better than she, Alice
30. When Janice discovered the school she was planning to enter was half Negro, she
31. Jane feels most people who meet her for the first time think she is
32. Betty thinks the things that make a person work hardest are
33. Dorothy believed most labor leaders are
34. As between mother's and father's ways Liz usually chose
35. When her favorite beauty shop began being used by Negroes, Barbara
36. Lillian feels most persons holding high public office are
37. Rosalind sometimes thought her father
38. Edith often felt like
39. When told she would have to do the whole thing by herself, Ruth
40. Peggy feels most people don't reach their goals because of
41. When Negroes began being admitted to the club, Ethel decided
42. Joan believes her family thinks she is
43. If Marge had any children, the one thing she would surely do for them is
44. When Nan thought the odds were against her, she
45. Dot would do anything in order to be
46. When the boss began hiring many Negroes, Evelyn
47. Estelle's feeling about most labor unions was that they
48. If Jean's father could change, Jean would want him to be
49. Mary believes most Russians want
50. From past experience, no matter what others thought, Peg herself knew that she
51. When they put a Negro to work next to her, Gail decided
52. Working with others all the time made Jane
53. When she was asked her opinion of public ownership, Kate said
54. Joan's fondest ambition was to become a
55. Every time they didn't invite Betty to the party, she
56. When Edith saw a Negro and a white involved in the accident, she naturally blamed the
57. When she found out the candidate belonged to a different religion, Ruth decided
58. Before the game started Alice felt her side would
59. If there is one group in the U. S. that can't be trusted, Isobel thought it was

THE PERSONAL AND SOCIAL ATTITUDES RECORD

For the use of schools, colleges, and records offices

Form III M-F

No.

Name

Age Sex Male Female

School or Office

Education

Date

The office would like to have in its files a record of your personal and social attitudes—that is, your attitudes toward politics, responsibility, parents, minority groups, race prejudice, etc.

Complete each of the statements by writing in what best applies in your case.

Consider each statement carefully. There is no time limit. *What you write in will represent the record of your personal and social attitudes.*

Do not skip any sentences.

1. When I think the job is too much for me, I tend to
2. I think most politicians are
3. If I heard I would have to join a union in a new job, I would
4. If I saw Negroes moving in next door, I would
5. When I am asked to be in charge, I generally tend to
6. The reason I try to get ahead is that I want to
7. If I saw the doctor they were trying to give me was a Negro, I would
8. Most jobs with responsibility make me feel
9. If I have something to say and others are around, I
10. If I saw my younger sister dancing with a Negro, I would—(If you don't have a younger sister, answer as if you did.)
11. If I were asked whether I could ever see myself voting Socialist, I would say:
12. After I finish a test I usually think that I have
13. If the man running for mayor admitted he did not believe in God, I would decide
14. I believe most people think of me as
15. I believe most labor leaders are
16. If I discovered the school I was planning to enter was half Negro, I would
17. As between mother's and father's ways, I choose
18. According to me, if a person remains poor it is because
19. If the boss began hiring many Negroes, I would
20. If I had any children, the one thing I would surely do for them is
21. I often feel like
22. When I am asked my opinion of public ownership, I say
23. If I see a Negro and a white involved in an accident, I naturally blame the
24. When I see others doing better than I
25. Every time they don't invite me to the party, I tend
26. If I found out the candidate belongs to a different religion, I would decide
27. If they put a Negro to work next to me, I would
28. Working with others all the time makes me
29. From past experience, I feel most people are
30. I feel most persons holding high public office are

Bureau Ednl. Res. Research

DAVID H. LANG COLLEGE

Dated

31. When I am told I have to do the whole thing by myself
32. My feeling about most labor unions is that they are
33. If my favorite beauty shop began being used by Negroes, I would
34. If they tried to get me to be like Mother, I would
35. If a new friend of mine said she was against God, I would
36. If they put a Negro forelady over me, I would
37. When I think the odds are against me, I tend to
38. I often think of myself as
39. If Negroes began being admitted to the club, I would decide
40. I sometimes think my father is

INSTRUMENTS USED IN STUDY II

Form B

THE G-W TEST OF VERBAL SPEED

Name Age Grade

Directions

This is a test to see how fast you can make sentences. But this is a different kind of test than you have taken before, for in this test there are no wrong answers. In the test you are given the first part of a sentence, and you are to write down the rest of the sentence. The best way to do this is to write down the *first thing* that comes to your mind when you read the beginning of the sentence. Don't worry about spelling, but write as quickly and as clearly as you can.

Example: When John scored a goal, he

You could finish this sentence by writing down: *was happy*, or *wanted to score another one*, or *stopped playing and went home*, or some other answer which makes sense.

Do Not Turn This Page Until Your Teacher Tells You to Begin.

- | | |
|---|---|
| 1. In his spare time Al likes to | 22. From past experience Harold thinks that most people |
| 2. When he went to the circus, Pete | 23. When Arthur breaks a window while playing, he |
| 3. When John's sister gets a bigger piece of cake, he | 24. When other people start telling him how to do something, Tim |
| 4. If the girls try to help Ben when he is hurt, he | 25. When the girls ask Lou to play on their team, Lou |
| 5. Chuck thinks that most bosses are | 26. If Stan saw Negroes moving in next door, he |
| 6. If Jack wasn't invited to the party, he | 27. When Arthur goes someplace with his whole family, he |
| 7. When the small boys have something which Eddie wants, he | 28. When he sees that others are doing better than he is, Roy |
| 8. When Bill is told to do something by himself, he | 29. When Paul has a chance to get even with the boy who cheated him, he |
| 9. Don sometimes thinks his family | 30. When another boy hits him, Tom |
| 10. When he was told that he failed the test, Bob | 31. When the captain of the team doesn't let Jim play, Jim |
| 11. If George had to work with girls all the time, he | 32. If they began letting Jewish boys into the club, Walter would |
| 12. When Joe thinks the homework is too hard for him, he | 33. When Charles has to go to the front of the class to explain something, he |
| 13. Whenever they ask Fred to be in charge, he | 34. Dick sometimes thinks his mother |
| 14. Henry sometimes thinks his father | 35. When Jerry thinks that the odds are against him, he |
| 15. Frank thinks that kids who always study hard | 36. If the teacher caught Ralph cheating on the test, Ralph |
| 16. When his brother gets a new toy, Roger | 37. When Harry sees a Catholic boy and a Protestant boy fighting, he blames |
| 17. When Ray finds himself alone with a group of important people, he | 38. When a girl asks Carl to help her, Carl |
| 18. If Mike found out that his new friend didn't believe in God, Mike | 39. When Ted can't do something the first time he tries it, he |
| 19. When they don't let him do what he wants, Ken | 40. When Dan sees that he is heading for a fight, he |
| 20. If Dave was sick and no one came to see him, he | |
| 21. When Steve has to talk to grown-ups he has never seen before, he | |

THE G-W TEST OF VERBAL SPEED

Name Age Grade

Directions

This is a test to see how fast you can make sentences. But this is a different kind of test than you have taken before, for in this test there are no wrong answers. In the test you are given the first part of a sentence, and you are to write down the rest of the sentence. The best way to do this is to write down the *first thing* that comes to your mind when you read the beginning of the sentence. Don't worry about spelling, but write as quickly and as clearly as you can.

Example: When Joan scored a goal, she

You could finish this sentence by writing down: *was happy*, or *wanted to score another one*, or *stopped playing and went home*, or some other answer which makes sense.

Do Not Turn This Page Until Your Teacher Tells You to Begin.

- | | |
|--|--|
| 1. In her spare time Thelma likes to | 22. From past experience Ann thinks that most people |
| 2. When she went to the circus, Patsy | 23. When Alice breaks a window while playing, she |
| 3. When Joan's brother gets a bigger piece of cake, she | 24. When other people start telling her how to do something, Carol |
| 4. If the boys try to help Betty when she is hurt, she | 25. When the boys ask Louise to play on their team, Louise |
| 5. Beverly thinks that most bosses are | 26. If Stella saw Negroes moving in next door, she |
| 6. If Janet wasn't invited to the party, she | 27. When Connie goes someplace with her whole family, she |
| 7. When the small girls have something which Eva wants, she | 28. When she sees that others are doing better than she is, Roberta |
| 8. When Judy is told to do something by herself, she | 29. When Paula has a chance to get even with the girl who cheated her, she |
| 9. Dot sometimes thinks her family | 30. When another girl hits her, Peggy |
| 10. When she was told that she failed the test, Donna | 31. When the captain of the team doesn't let Helen play, Helen |
| 11. If Ruth had to work with boys all the time, she | 32. If they began letting Jewish girls into the club, Wilma would |
| 12. When Elaine thinks the homework is too hard for her, she | 33. When Irene has to go to the front of the class to explain something, she |
| 13. Whenever they ask Freda to be in charge, she | 34. Jean sometimes thinks her mother |
| 14. Gloria sometimes thinks her father | 35. When Gerry thinks that the odds are against her, she |
| 15. Frances thinks that kids who always study hard | 36. If the teacher caught Bea cheating on the test, Bea |
| 16. When her sister gets a new toy, Rose | 37. When Barbara sees a Catholic girl and a Protestant girl fighting, she blames |
| 17. When Agnes finds herself alone with a group of important people, she | 38. When a boy asks Marie to help him, Marie |
| 18. If Mary found out that her new friend didn't believe in God, Mary | 39. When Donna can't do something the first time she tries it, she |
| 19. When they don't let her do what she wants, Kay | 40. When Liz sees that she is heading for a fight, she |
| 20. If Nancy was sick and no one came to see her, she | |
| 21. When Susan has to talk to grown-ups she has never seen before, she | |

THE G-W TEST FOR BOYS AND GIRLS

Name Age Grade

Directions

This is a test to find out how you behave when certain things happen to you. In this test you are given the first part of a sentence, and you are to write down the rest of the sentence. There are no wrong answers in this test, but your answers will tell something about what kind of a boy or girl you are. Don't worry about spelling, and don't waste time. Write as clearly as you can.

Example: When my sister wins, I

One child might write down: *am very happy*. Another child might write: *become angry*. Or a third child might answer: *say she cheated*. The answer which you write should tell what you do when your sister wins.

If you do not have any sisters or brothers, just pretend that you have them when you come to a sentence which talks about brothers and sisters.

Do Not Turn This Page Until Your Teacher Tells You to Begin.

- | | |
|--|---|
| 1. In my spare time I like to | 22. From past experience I think that most people |
| 2. When I went to the circus, I | 23. When I break a window while playing, I |
| 3. When my sister gets a bigger piece of cake, I | 24. When other people start telling me how to do something, I |
| 4. If the girls try to help me when I am hurt, I | 25. When the girls ask me to play on their team, I |
| 5. I think that most bosses are | 26. If I saw Negroes moving in next door, I |
| 6. If I wasn't invited to the party, I | 27. When I go someplace with my whole family, I |
| 7. When the small boys have something which I want, I | 28. When I see that others are doing better than I am, I |
| 8. When I am told to do something by myself, I | 29. When I have a chance to get even with the boy who cheated me, I |
| 9. I sometimes think my family | 30. When another boy hits me, I |
| 10. When I was told that I failed the test, I | 31. When the captain of the team doesn't let me play, I |
| 11. If I had to work with girls all the time, I | 32. If they began letting Jewish boys into the club, I would |
| 12. When I think the homework is too hard for me, I | 33. When I have to go to the front of the class to explain something, I |
| 13. Whenever they ask me to be in charge, I | 34. I sometimes think my mother |
| 14. I sometimes think my father | 35. When I think that the odds are against me, I |
| 15. I think that kids who always study hard | 36. If the teacher caught me cheating on the test, I |
| 16. When my brother gets a new toy, I | 37. When I see a Catholic boy and a Protestant boy fighting, I blame |
| 17. When I find myself alone with a group of important people, I | 38. When a girl asks me to help her, I |
| 18. If I found out that my new friend didn't believe in God, I | 39. When I can't do something the first time I try it, I |
| 19. When they don't let me do what I want, I | 40. When I see that I am heading for a fight, I |
| 20. If I was sick and no one came to see me, I | |
| 21. When I have to talk to grown-ups I have never seen before, I | |

THE G-W TEST FOR BOYS AND GIRLS

Name Age Grade

Directions

This is a test to find out how you behave when certain things happen to you. In this test you are given the first part of a sentence, and you are to write down the rest of the sentence. There are no wrong answers in this test, but your answers will tell something about what kind of a boy or girl you are. Don't worry about spelling, and don't waste time. Write as clearly as you can.

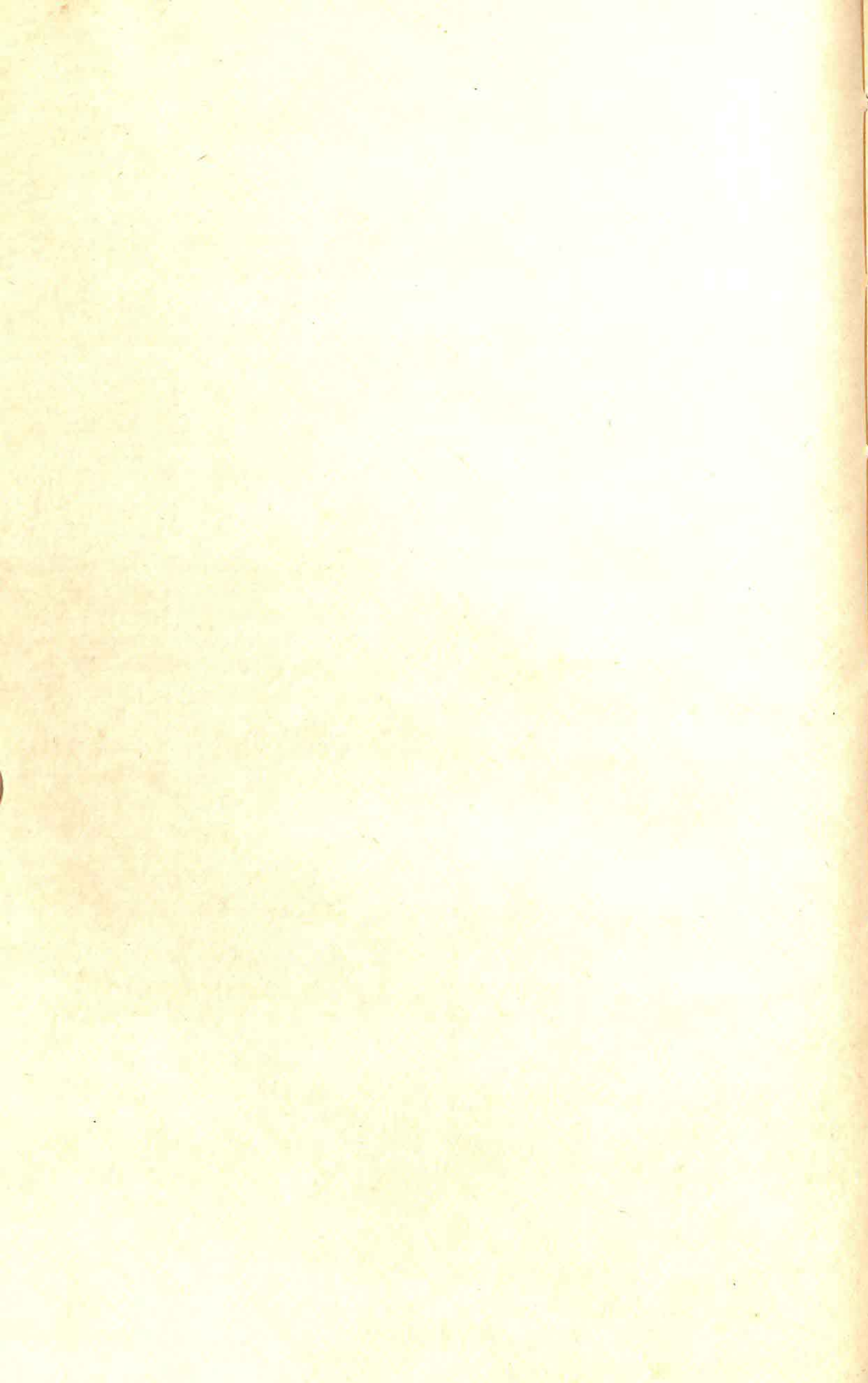
Example: When my sister, wins, I

One child might write down: *am very happy*. Another child might write: *become angry*. Or a third child might answer: *say she cheated*. The answer which you write should tell what you do when your sister wins.

If you do not have any sisters or brothers, just pretend that you have them when you come to a sentence which talks about brothers and sisters.

Do Not Turn This Page Until Your Teacher Tells You to Begin.

- | | |
|--|---|
| 1. In my spare time I like to | 22. From past experience I think that most people |
| 2. When I went to the circus, I | 23. When I break a window while playing, I |
| 3. When my brother gets a bigger piece of cake, I | 24. When other people start telling me how to do something, I |
| 4. If the boys try to help me when I am hurt, I | 25. When the boys ask me to play on their team, I |
| 5. I think that most bosses are | 26. If I saw Negroes moving in next door, I |
| 6. If I wasn't invited to the party, I | 27. When I go someplace with my whole family, I |
| 7. When the small girls have something which I want, I | 28. When I see that others are doing better than I am, I |
| 8. When I am told to do something by myself, I | 29. When I have a chance to get even with the girl who cheated me, I |
| 9. I sometimes think my family | 30. When another girl hits me, I |
| 10. When I was told that I failed the test, I | 31. When the captain of the team doesn't let me play, I |
| 11. If I had to work with boys all the time, I | 32. If they began letting Jewish girls into the club, I would |
| 12. When I think the homework is too hard for me, I | 33. When I have to go to the front of the class to explain something, I |
| 13. Whenever they ask me to be in charge, I | 34. I sometimes think my mother |
| 14. I sometimes think my father | 35. When I think that the odds are against me, I |
| 15. I think that kids who always study hard | 36. If the teacher caught me cheating on the test, I |
| 16. When my sister gets a new toy, I | 37. When I see a Catholic girl and a Protestant girl fighting, I blame |
| 17. When I find myself alone with a group of important people, I | 38. When a boy asks me to help him, I |
| 18. If I found out that my new friend didn't believe in God, I | 39. When I can't do something the first time I try it, I |
| 19. When they don't let me do what I want, I | 40. When I see that I am heading for a fight, I |
| 20. If I was sick and no one came to see me, I | |
| 21. When I have to talk to grown-ups I have never seen before, I | |



Psychological Monographs: General and Applied

The Use of Rotter's Social Learning Theory in Developing a Personality Inventory^{1,2}

SHEPHARD LIVERANT

Ohio State University

THIS RESEARCH represents an effort to design and validate a psychometric measure of a theoretical construct derived from a formalized theory of behavior. The model provided by Rotter's (14) Social Learning Theory of Personality (SLT) has been employed in this project.

Several authors (1, 4, 8, 17) have argued for a more "rational approach" in the design of personality tests; but, as Butler points out:

"... the writer knows of no personality inventory for which the content, the form of items and the psychometric methods applied have been dictated by a formal psychological model" (1, p. 77).

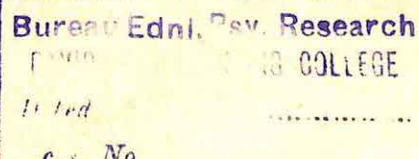
A survey of the literature supports Butler's contention by indicating that the great majority of attempts to improve test construction has been devoted to de-

veloping more adequate mathematical models for psychological tests which adhere to the tenets of physical measurement. Whereas such efforts are essential, even the most refined mathematical model is not a sufficient means of devising adequate measuring instruments for personality constructs. It is clear that the mathematical model alone cannot yield test items, but can only provide mathematical procedures to justify their final inclusion in a test. In addition to the difficulties involved in constructing appropriate items, another major source of potential failure in devising a useful measuring instrument may be the nature of the construct being measured. Some psychological constructs are more useful tools for explaining behavior than others. For example, the construct of intelligence is considered superior in this respect to that of honesty. Test construction based primarily on a mathematical model does not provide any means of establishing the value of the construct being measured until after the test has been validated. If the test emerges as a valid and useful instrument, the construct was probably a good one. However, if the test proves invalid, then either the construct may have been poor or the instrument was inappropriately designed to measure this construct. Which factor is operating cannot be readily determined.

To a large extent, the selection of con-

¹This study is a condensation of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Colorado, June, 1956.

²The writer wishes to express his indebtedness to Richard Jessor whose continued support, both academic and personal, enabled the completion of this study. Of the many who contributed to this project the assistance of Maurice P. Smith, Victor C. Raimy, Kenneth R. Hammond, and William B. Sawrey is gratefully acknowledged. A special vote of thanks is owed to the large number of undergraduate students who so patiently and willingly took the various forms of this test. Finally, I wish to express my appreciation to Jo Ann Zaynor, Carolyn Harris, Darlene Thomas, Gladys Sensbach, and Marie Richardson (now my wife) for their statistical and editorial assistance.



structs and items has depended upon the implicit preferences of the personality test constructor. The use of an explicit methodology concerning their selection may prove to be a more fruitful approach to objective test construction. Behavior theories provide a more explicit and controlled means of selection of both the constructs to be measured and the items used in measuring them. An adequate mathematical model coordinated with a meaningful psychological theory may prove to be both the necessary and sufficient conditions for the development of satisfactory personality inventories.

The use of behavior theory in test construction can be of great potential benefit to both theory and test building. The construction of instruments to measure theoretical constructs should enable more adequate verification or refutation of hypotheses deduced from the system and thus lead to a more valid and powerful conceptualization of the theory. This in turn should provide a means of constructing measures of greater validity. It is the continuation of this cycle which may eventually provide a greater understanding of any set of behaviors taken under consideration.

One aim of this study is to provide empirical justification for the behavior theory approach to test construction. The other is to develop an instrument which can be used by subsequent investigators interested in further development of the theory from which the test was derived. It is the accomplishment of the latter aim which provides crucial support for the value of the behavior theory approach to test construction. The value of this approach can be considered partially demonstrated when it fulfills these essential conditions: the items used to formulate the measuring instrument

should be selected in a manner proposed by the theory; the test should be psychometrically sound and measure an explicitly stated theoretical construct; and, lastly, the test should be useful in testing experimental hypotheses deduced from this same theory.

SOCIAL LEARNING THEORY

There are several reasons for the selection of SLT for test construction purposes. Rotter has placed a great deal of emphasis on the possible interrelationship of four classes of variables. These include the subject's measurable behavior, his expectations that his behavior will be followed by particular kinds of events or reinforcements, the value of these external events or reinforcements, and the psychological situation in which behavior takes place. Knowledge of the interplay of these classes of variables has indispensable guiding value for the personality test constructor. Perhaps most important is the fact that the constructs are given operational definition, which allows the selection of clearly specified behavioral referents so necessary for the valid measurement of psychological variables. Aspects of the theory relevant to the present problem will be considered briefly in the following section.

In SLT, *need* is the abstraction from learned behaviors which have common directionality. It is hypothesized that various behaviors become functionally related (lead to the same or equivalent goals) through a process of learning and generalization. In this system goals are functionally related if they evoke the same or similar behaviors, and behaviors which lead to the same or similar goals form a functional class. This viewpoint of Rotter's has resulted in a working categorization of learned behaviors into

six major needs: recognition-status, protection-dependency, dominance, independence, love and affection, and physical comfort.

Each of these need categories represents what Peak (13) refers to as a functional unity. To say that behaviors are functionally related or organized means, in this case, that they change concomitantly. An event which affects one of the behaviors in the unit or class will affect all other behaviors in that unit in a predictable manner. In other words the behaviors within a functional class are correlated. The notion of concomitant variation based on functional relationships allows the prediction of occurrence of behavior X as a result of measuring or observing the class of functionally related behaviors (need) to which X belongs.

The construct "need" describes a person's tendency to behave in a given fashion in certain situations. The mean potentiality of a group of functionally related behaviors occurring in any situation is what is meant by strength of need, or, to use Rotter's term, "need potential." Strength of need depends upon the interaction of the variables *freedom of movement* and *need value*. The mean expectancy (subjective probability) that an individual has of obtaining positive reinforcement as a result of employing a class of functionally related behaviors constituting a particular need is known as the freedom of movement for that need. The mean preference value for the reinforcements expected by employing a class of functionally related behaviors is what is meant by need value. To arrive at an estimate of the strength of need, some type of measurement of freedom of movement and need value is necessary. Consequently, the construc-

tion of reliable and valid measures to serve as indicants of freedom of movement and need value is a necessary step in estimating the predictive value of these constructs and, thus, their explanatory validity.

The construct "need value," because it is hypothesized to be less subject to immediate environmental influence than freedom of movement, is more amenable to psychometric measurement at the present stage of the theory's development. Need value pertains to the preference for one set of reinforcements over another when expectancy of attaining either set is held constant. The mean preference value of a set of functionally related reinforcements relative to other sets is what is meant by need value.

As is obvious from the role of this construct in Rotter's theory, actual behavior cannot be predicted from a measure of need value alone. It is, however, one of the major variables which must be accounted for in predicting behavior from this theoretical framework. The interaction of other classes of variables with each other and with need value must be known or controlled to allow for actual behavioral prediction. This point seems to have been neglected frequently in the interpretation of other measuring instruments.

The basic situation for the measurement of need value is simply this: Present a subject with two or more sets of functionally related behaviors (i.e., the members of each set have been demonstrated by some empirical means to lead to the same or similar goals for that subject). Require the subject to make statements of preference or choice between two behaviors from each set, paired in every possible combination, with the stated provision that he has equal oppor-

tunity for engaging in either (thereby controlling expectancy). The number of items picked in one set as most preferable represents an indication of need value for that series of behaviors (need) as compared to the set or sets with which it was contrasted.

This operational definition of need value makes some type of choice or ranking technique mandatory for developing an indicant of this construct. On the basis of this, and of certain psychometric considerations derived from previous research (3, 9), a forced choice technique was constructed to measure need value.

By theoretical definition, the measurement of need value requires that expectancy be held constant. The control of expectancy was attempted via the use of instructions. Each choice is presented to the subject with the instructions: "Select the one you most prefer to do if you had an *equal* opportunity to do either one." The effectiveness of such instructions for controlling expectancy has not been tested and may be considered, therefore, a weakness of the design.

Not only did Rotter's theory dictate the form of the item but the content as well. The content of the items designed to measure a particular need value is determined by sampling a population of behaviors constituting a given need category. Since the measurement of a particular need value is relative to other known sets of behaviors, at least two need categories must be sampled.

The breadth of Rotter's classification scheme provides a great variety of behaviors about which predictions can be made but, consequently, lowers the accuracy of prediction regarding any specific event. The level of generality at which these needs are abstracted transcends all situations. Throughout his

theory, Rotter emphasizes the importance of the psychological situation as a determiner of behavior. Lowering the level of generality of these needs by sub-categorizing them in terms of specific situations should significantly increase accuracy of prediction. This reasoning resulted in selecting items which fitted four situationally qualified need categories.

Rotter cautions that a need category based on functionally related behaviors must be determined empirically and that the classification or organization of behaviors into psychological needs must be considered characteristic only of the particular culture in which the classification is made. Behaviors which bring love, recognition, or comfort in one culture are not necessarily those which bring the same reinforcements in another culture.

The latter two statements suggest that clues to some functional relationships are given by the culture itself. Judges familiar with the college culture were asked to list behaviors which they thought individuals in a college population use to satisfy the four need categories under consideration. Those behaviors having high interjudge agreement were used to make up the test items.

It is realized that cultural determination of need categories does not take into account the individual deviations of the members of that culture. Any individuals employing behaviors different from those of the majority of people in their cultural group in an attempt to attain the same goals will not be taken into account by a measure constructed in this manner. Such individual deviations will constitute error variance of a systematic nature in this type of scale. However, to arrive at a definitive appraisal of any

one individual's need hierarchy requires elaborate measurement techniques which, as Rotter suggests:

"... are neither economical nor feasible for the study of groups or in collection of group data in order to test theoretical hypotheses. For these latter purposes the method ... using cultural referents would appear to be adequate" (14, p. 186).

Operationally, there is a significantly higher correlation in the way in which people react to behavioral referents within a need than between needs. The effectiveness of the judgmental method for establishing functional unities or needs on a cultural basis was tested by means of an item analysis and a factor analysis.

TEST CONSTRUCTION

Category Selection

The operations involved in developing a need category provide a basis for grouping behavioral items into scorable categories. The names of the needs and their definitions are listed below:

1. Recognition-Status in Academic Situations (AcR): Need to be considered competent or good in academic situations. The need to have your academic behaviors approved and admired by others. Need to gain academic status.

2. Recognition-Status in Social Situations (SoR): Need to be considered competent or good in social activities. The need to have your social behaviors approved and admired by others. The need to gain social status.

3. Love and Affection in Academic Situations (ALA): Need for acceptance and indications of liking by your classmates in academic situations. Need to be valued as a friend in academic situations.

4. Love and Affection in Social Situations (SLA): Need for acceptance and indications of liking by other individuals in a social atmosphere. Need to feel the

sheer joy of being with others regardless of any advantage to yourself. Need to feel part of a social group, i.e., to be valued as a friend in social activities.

These needs were selected for the following reasons:

1. The behaviors subsumed under each need are familiar, through personal experience and daily contact, to the author and to the judges used in this study.

2. The behaviors subsumed under each need are likely to differentiate students in terms of frequency of use. If most members of the population employ the referent behaviors to approximately the same extent, these behaviors will not be useful for constructing test items designed to differentiate subjects in terms of strength of need.

3. These needs refer to classes of behavior which often seem to come into conflict in the college setting. The selection of a behavior subsumed under one need often is incompatible with engaging in a behavior subsumed under another need. An attempt was made to approach a more representative sample of behavioral choices by making the test choices as "lifelike" as possible. In the case of many items this criterion could not be met because it came into conflict with other criteria used to pair the behaviors, namely, social desirability and item intercorrelation.

4. In the selection of the four needs an attempt was made to reach a suitable compromise between breadth and exactitude of prediction. The two major need areas, recognition and love and affection, were subdivided in terms of situations likely to be encountered by college students.

Behavior Selection

A dittoed pamphlet consisting of a

brief orientation to Rotter's conceptualization of needs, the four needs followed by their logical definitions, and six illustrative behavioral referents was constructed by the author. These pamphlets were distributed to faculty members and graduate students in the Department of Psychology at the University of Colorado. They were instructed to list 10 or more behaviors which they thought the majority of college students use to satisfy each need.

The submitted behaviors were sorted to prevent duplication and edited for clarity of expression. They were then pooled and judged by four judges (including the author) as to their relevance as referents for each need. This was accomplished by means of an item sort. Each judge was presented all the behavioral referents transcribed on 3-by-5 cards and asked to sort them into four piles corresponding to the four needs. The judges were instructed to separate out those behaviors which seemed to overlap more than one category or for any other reason did not seem well suited to any one of the four needs. Only those behaviors which at least three out of the four judges agreed upon were considered in the construction of the items. The probability that three out of four judges will select the same referent to fit the same need strictly on the basis of chance is 1 in 64. This means that out of the 239 submitted behavioral referents only four would be likely to be selected by chance alone.

Preliminary Scale Construction

Each item in the scale is formed by pairing a behavioral referent from one need category with a behavioral referent from another need category. A crude attempt to control social desirability was

made by pairing referents which had approximately equivalent social acceptance value in terms of the author's judgments.

Combining referents from each need category with referents for each of the other three needs resulted in the construction of six subtests, each one including 20 items. The subtests and the contrasting need categories are listed below:

Subtest	Needs
I.	AcR vs. SoR
II.	AcR vs. ALA
III.	AcR vs. SLA
IV.	SoR vs. ALA
V.	SoR vs. SLA
VI.	ALA vs. SLA

Scoring

The total score in any one of the six subtests is the sum of the alternatives, within a need category, which are selected by the subjects. Six scoring keys, one for each subtest, were designed to enable a count of the number of alternatives checked in the first-named need of each subtest. For example, a score of 15 on subtest I (AcR vs. SoR) means that the individual has selected 15 alternatives referring to academic recognition and 5 referring to social recognition. The same scheme is followed in determining a score for each of the other five subtests.

The score on one need category within a subtest is, of course, the complement of the score on the need with which it is contrasted. To say that a subject has a score of 15 on subtest I means that he has a score of 15 on the need for academic recognition and a score of 5 on social recognition, provided he has answered all the questions. The scores between subtests, however, are mechanically independent. The format of the test in no way makes the score on one subtest de-

pendent on the scores made on any of the other subtests.

The higher the score on a particular need category, the more often the subject has chosen the statements for this need as being more preferable activities than the statements for the contrasting need. The lower the score, the less often the subject has chosen them as being more preferable. By definition, then, the total score on any subtest is an indicator of strength of need value in terms of frequency of choice for one need relative to another need. In this form of the inventory no attempt was made to combine subtest scores.

The trial form was pretested on 22 students in an introductory psychology class. There were no difficulties connected with the instructions and the form proved to be a very convenient one to administer and score. It requires about 25 minutes for the average college student to complete the test, and almost all college students will finish by the end of 50 minutes.

The Development of the Second Form

As a test of the adequacy of the preliminary form and to provide a basis for building a new form, scores on each of the six subtests were obtained from 125 students (70 males and 55 females) in another introductory psychology class. For the theoretical and psychometric reasons discussed in the previous chapter, the kind of items wanted in this test were those which intercorrelated within a subtest and which are relatively immune to being answered on socially desirable grounds.

An item analysis served as a check on our judgments in selecting related items to compose the need categories. A biserial r was computed for each item to find

the relationship between each item and the total subtest score. Each item was dichotomized in terms of the alternatives a or b and correlated with the total subtest score. In essence this is a rough device used in the initial stages as an aid in establishing the existence of functional unities within our test. Any item which did not demonstrate a degree of correlation with the total subtest score at or greater than the 5% level of confidence was considered invalid.

An estimate of social desirability was obtained for each item by calculating the proportion of subjects selecting alternative a as compared to those selecting alternative b in the case of each item. If more than 80% of the sample taking the trial form selected either alternative, then that item was judged as being responded to in terms of a social desirability factor.

Any item which did not meet both criteria, i.e., demonstrating a significant correlation with the total subtest score and having neither alternative selected by more than 80% of the subjects was discarded. Forty-four out of the original 120 were scrapped or reassembled on this basis. Of the remaining 76 items, 12 more had to be recombined to enable the use of the same 20 behavioral referents for each of the four needs. The remaining 64 items formed a nucleus around which a new form of the test was built.

In the pairing of different need alternatives to make up new items, a further attempt to control social desirability was made. The "social desirability value" of each alternative was obtained by counting the number of times it was selected in each of its pairings in the preliminary form (ranging from one to three) and computing the mean. Referents from each need were paired with referents

from each of the other three needs for approximate equality of social desirability as defined by this kind of frequency count.

The finding of Gordon (10) and Edwards (5) that there is a high positive relationship between independently obtained social desirability indices and frequency of choice lends some justification to the use of this method for controlling social desirability. However, the method used in this study must be considered inadequate because, for one thing, the estimate of social desirability is based on the number of times an alternative is selected after it is put in a forced-choice item to which the subjects are instructed to respond on a need value basis. Consequently, it is very likely that the social desirability factor and the variable being measured are contaminating each other and there is no way of accurately determining the influence of either one separately. The superiority of Edwards' (6) scaling method for controlling social desirability is unquestioned, but the amount of time and effort required prevented its use in the present study.

In order to permit totaling of subtest scores, the *same* 20 behavior statements from each need were paired with the *same* 20 behavior statements from each of the other three needs. The format and instructions of the second inventory are the same as that of the trial form. The same six subtests are composed of 20 forced-choice pairs each, which makes a total of 120 items. For ease in scoring, the items on the inventory are grouped in blocks of six. Each block has one item from each subtest arranged in consecutive order.

Scoring Combinations

The subtests are scored by the same

method as described for the trial form. The repeated use of the 20 referents from each category provides the rationale for combining subtest scores in various ways. The scores used in this study are listed below:

Subtest	Need Categories	Range of Scores
I	AcR vs. SoR	0-20
II	AcR vs. ALA	0-20
III	AcR vs. SLA	0-20
IV	SoR vs. ALA	0-20
V	SoR vs. SLA	0-20
VI	ALA vs. SLA	0-20

Combination of Subtest Scores Yielding Situational Need Scores

	Range of Scores
Subtest I + subtest II + subtest III = Need for Academic Recognition	0-60
*Complement of Score of Subtest I + Score of Subtest IV + Score of Subtest V = Need for Social Recognition	0-60
Complement of Score of Subtest II + Complement of Score of Subtest IV + Score of Subtest VI = Need for Academic Love and Affection	0-60
Complement of Score of Subtest III + Complement of Score of Subtest V + Complement of Score of Subtest VI = Need for Social Love and Affection	0-60

Over-All Need Scores

Score on Subtest II + III + IV + V =
Need for Recognition
Complementary Scores on II + III + IV
+ V = Need for Love and Affection
Range of both Need Scores is 0-80

The scoring arrangements presented above represent three levels of abstraction.

* Formula for obtaining complementary scores:
20 - Subtest Score = Complementary Score
Example: 20 - Score on Subtest I (AcR Score)
= SoR Score

tion, each one more inclusive than the preceding one. The scoring method to be used in describing a person's need values depends on the breadth and/or exactitude of prediction desired. The utility of various scoring combinations for predicting specific behaviors will determine the preference accorded them.

The second form of the test was administered to a different group of subjects from a course in general psychology ($N = 148$). Once again, a biserial r was computed for each item to determine its relationship with the total subtest score. Another frequency count of the proportion of subjects answering either alternative of each item was made. The results of these two procedures led to the decision that this form of the inventory was satisfactory for further investigative procedures. Only eight (items 14, 111, 58, 12, 30, 60, 84, 108) were not significant at the 5% level of confidence and five of these, the last numbered, are in subtest VI. In 10 cases (items 32, 44, 46, 111, 28, 41, 30, 102, 108, 114) more than 80% of the subjects selected either alternative. The procedures to be discussed in the following chapters are based on this form of inventory, named the Goal Preference Inventory (GPI), which is presented in its entirety in the Appendix.

EVALUATION OF THE GPI³

Norms

The norms available at the present time are for unmarried, undergraduate students at the University of Colorado who at the time of testing were enrolled

³ This evaluation has been heavily influenced by the "construct validity" approach described by Cronbach and Meehl (4). The present project (i.e., the use of theory in test construction) is viewed as a logical extension of their point of view.

in introductory psychology classes. The normative sample consists of 233 women and 157 men ranging in age from 18 to 32 years. This population (composed largely of sophomores) is the group ordinarily used for research purposes by psychologists at the university. All subjects in the normative group were told that they were taking this test for research purposes.

Distributions of scores on the subtests were made separately for men and women to determine the presence of sex differences. The means and standard deviations for each of the subtest scores and for each of the scoring combinations are presented in Table 1. Critical ratios were computed to test the significance of difference between the mean score of men and the mean score of women in each scoring category.

The raw score distributions for both men and women appear to form essentially normal curves in the case of each scoring category. What constitutes a high or low score on any distribution is best determined by each user of the test in terms of his own objectives. The means and standard deviations shown in Table 1 provide a convenient basis for converting raw scores into standard scores for the individuals in the normative sample. Although it was not done in this study, the subtest scores can be converted into standard scores and then summed to arrive at the need scores.

Internal Consistency

The test was constructed with the aim of having the items within each subtest measure the same attribute or variable. The computation of split-half reliability coefficients for each of the subtests indicates how well we have approached this goal. Odd-even scores were correlated for

TABLE 1
MEANS^a AND STANDARD DEVIATIONS OF THE RAW SCORE DISTRIBUTIONS FOR THE
NORMATIVE SAMPLE

Means				Standard Deviations		
Variable	Men	Women	Total	Men	Women	Total
Subtest I	10.84**	10.04	10.37	3.82	3.49	3.64
II	11.03**	10.23	10.55	3.30	2.85	3.06
III	9.97*	7.87	8.72	3.48	3.13	3.44
IV	10.19	10.20	10.20	3.73	3.28	3.47
V	9.28*	8.16	8.61	3.83	3.53	3.69
VI	9.25*	8.02	8.52	2.32	2.26	2.36
AcR	31.71*	28.14	29.58	8.07	7.35	7.85
SoR	28.76	28.34	28.51	9.85	8.58	9.12
ALA	28.05	27.64	27.80	5.97	5.21	5.54
SLA	31.48	35.98*	34.17	7.00	6.69	7.16
Over-all Rec.	40.55*	36.50	38.13	10.82	9.60	10.30
Over-all L & A	39.45	43.50*	41.87	10.82	9.60	10.30
N	157	233	390			

^a The mean is significantly larger than the corresponding mean for the opposite sex.

* Significant at the .01 level.

** Significant at the .05 level.

each subtest over the 148 subjects in class one. The internal consistency coefficients, corrected by the Spearman-Brown formula, are given in Table 2.

As shown in Table 2, the range of coefficients is from 0.51 to 0.79 with the mean being 0.66. Taking into consideration the relative shortness of each subtest these coefficients indicate a reason-

ably high degree of internal consistency for the first five subtests. For example, they compare favorably with the internal consistency coefficients obtained by Edwards (7) with the Personal Preference Schedule which range from .60 to .87 with a mean of .76.

Stability

Stability coefficients are based upon a group of 62 students who took the test twice with a five-day interval separating the two administrations. These coefficients and the corresponding standard errors of measurement are given in Table 3.

In general, the results indicate a battery of subtests, excluding subtest VI, which is reliable for general predictive purposes, particularly when scoring combinations are used. The range is from .56 to .92 with a mean of .82. The size of the standard errors of measurement indicates that the absolute errors of measurement of the obtained scores are relatively small.

TABLE 2
COEFFICIENTS OF INTERNAL CONSISTENCY
FOR THE SIX SUBTEST SCORES

Subtests	Internal Consistency ^a	Mean ^b	SD
I	.73	10.41	3.78
II	.63	10.73	3.30
III	.62	9.12	3.55
IV	.70	10.27	3.71
V	.79	9.10	4.41
VI	.51	8.58	2.73
N	148		

^a Split-half is based on odd against even items within each subtest corrected by the Spearman-Brown formula.

^b The means and standard deviations for each variable are based on the total raw scores for the 148 subjects.

As in the case of internal consistency, subtest VI yields a considerably lower stability coefficient than any of the other scoring categories. It is probable that all the scoring combinations incorporating subtest VI have lower stability coefficients as a consequence of the low reliability of subtest VI.

The relatively lower split-half and test-retest reliability coefficients of subtest VI (ALA vs. SLA) may indicate that the majority of subjects does not find it meaningful to choose between academic and social love and affection statements. As a result of the ambiguity involved in forcing such a choice, there may be a lack of response consistency within subtest VI at a given time and fluctuations in preference over a period of time. The following factor analysis tends to support this conclusion.

Factorial Validity

By the nature of its construction, the present test purports to measure four factors, i.e., the four need categories labeled AcR, SoR, ALA, and SLA. In an attempt to find the number and kinds of factors being measured by this test, the six subtests were intercorrelated and the resulting matrix was factor analyzed. The factor analysis in this case is used as an estimate of response consistencies. It is

TABLE 3

COEFFICIENTS OF STABILITY^a AND STANDARD ERRORS OF MEASUREMENT FOR ALL SCORING CATEGORIES

Scoring Categories	Stability			
	<i>r</i>	Mean	<i>SD</i>	<i>SE</i>
Subtest I	.90	10.74	4.32	1.38
II	.77	10.48	3.20	1.58
III	.86	8.44	3.81	1.45
IV	.86	9.58	3.74	1.42
V	.80	8.00	3.74	1.68
VI	.56	8.39	2.20	1.45
AcR	.91	20.66	9.26	2.78
SoR	.92	26.81	10.47	2.93
ALA	.80	28.32	5.81	2.61
SLA	.78	35.24	6.50	3.06
Over-all Rec.	.86	36.52	10.13	3.85
Over-all L & A	.86	43.48	10.13	3.85
<i>N</i>	62			

^a Test-retest with a five-day interval. Means and standard deviations are for first testing.

another test of the hypothesis that people will react to behavioral referent statements *within* a need category in a more similar fashion than they will to behavioral referent statements *between* need categories, i.e., factor analysis furnishes further evidence for the existence, or lack of existence, of functional unities.

Thurstone's (16) centroid method was used to factor analyze the subtest intercorrelation matrix shown in Table 4. This method enabled the extraction of

TABLE 4
INTERCORRELATIONS OF THE SUBTEST SCORES

Subtest ^a (<i>N</i> = 148)	II	III	IV	V	VI
	AcR vs. ALA	AcR vs. SLA	SoR vs. ALA	SoR vs. SLA	ALA vs. SLA
I. AcR vs. SoR		.28	-.60	-.43	.14
II. AcR vs. ALA	.25	.63	.29	.37	.025
III. AcR vs. SLA			.26	.48	.41
IV. SoR vs. ALA				.72	-.073
V. SoR vs. SLA					.22
VI. ALA vs. SLA					

^a Means and standard deviations for each subtest appear in Table 2.

TABLE 5
CENTROID-FACTOR MATRIX

Subtest	Factor Loadings			
	I	II	III	Communalities
I	-.30	.78	.10	.66
II	.50	.42	.45	.64
III	.68	.64	.00	.81
IV	.80	-.47	.25	.93
V	.89	-.14	-.17	.82
VI	.23	.32	-.39	.33

the three factors shown in Table 5. The third factor residuals were so small (largest coefficient was 0.05) that it was clear that there was nothing left but error variance. Orthogonal rotation of the reference axes provided the final rotated matrix shown in Table 6.

An examination of the factor structure shown in Table 6 reveals the presence of three factors. There are three subtests which have substantial loadings in common in each factor in contrast with the three other subtests that have zero or low factor loadings in each of the three factors. The common feature of the highly factor-loaded group of tests that is not shared by the other tests is used to define or interpret each of the three factors.

Tests I (AcR vs. SoR), IV (SoR vs. ALA) and V (SoR vs. SLA) have high loadings on Factor A. The high negative factor loading for subtest I means that

TABLE 6
ROTATED ORTHOGONAL FACTOR MATRIX

Subtests	Factors		
	A	B	C
I	-.68		
II	.21	.12	.47
III	.17	.02	.76
IV	.95	.55	.73
V	.78	-.07	.15
VI	-.04	.45	.18
		.54	.10

this test correlated positively with subtests IV and V in terms of the social recognition score, which it will be remembered is the complement of the academic recognition score on subtest I. The common feature of these three subtests, then, is social recognition, and Factor A is identifiable as a social recognition factor.

Factor B appears to be a social love and affection factor. It is defined by subtests III, V, and VI, whose common feature is social love and affection.

Factor C is defined by subtests I, II, and III, whose common feature is academic recognition.

Thus, the three factors which emerge can be identified as constituting three of the need categories used in constructing this test. The emergence of these three factors tends to support, among other things, the method of item selection via judges' item sorts. The factorial results also increase our confidence in the use of cultural cues as a means of grouping functionally related behaviors into relatively independent need categories.

The factor analysis further indicates that within the normative group the academic love and affection statements do not reflect a functional unity and when paired with social love and affection statements serve to lower the internal consistency and reliability of the scale. Consequently, the inventory could be improved by eliminating the academic love and affection category or substituting a more meaningful category.

The discovery of the lack of existence of an academic love and affection factor re-emphasized the importance of taking into account the meaning of the situation for the individual or class of individuals being studied. The academic love and affection factor was put into the test in

order to arrive at a symmetrical inventory, that is to say, as a balance for social love and affection in the manner that academic recognition balances social recognition. It proved the most difficult category for which to get referents, and some members of the original referent providing group claimed that academic love and affection does not exist. The need was established by psychology graduate students and staff, who probably perceive the classroom situation somewhat differently than undergraduate students. Whereas the test constructors may consider it meaningful and important to engage in classroom behaviors which lead toward a goal of love and affection, a large number of people in the tested group probably did not. It would seem advisable in future projects of this nature to have the behavioral referents for various need categories selected by a group which is representative of the population to be studied. The representativeness of the judges becomes especially important when attempting to establish tenuous or not widely accepted need categories such as academic love and affection.

Known Groups

It was emphasized in the previous chapter that a measure of need value does not alone enable accurate prediction of a given individual's behavior. Other factors have to be taken into account, a major one being expectancy. However, if these other factors are allowed to vary at random within groups, a coarse correspondence should be found between certain need value scores and designated group differences. For example, knowledge of the ways of life of two groups of people may lead us to expect that one group would have a higher need

value for certain goals than the other group. If our expectation is verified in terms of differential test scores, then both the construct and the test are validated to a certain degree.

In no case of known group validity used in this study is the differentiating group variable the direct criterion of need value, for the test is meant to measure something other than fraternity affiliation, etc. Only a *low* degree of correspondence is expected between test scores and designated groups because, as the result of the operation of numerous uncontrolled variables, members of the groups should overlap on the test. An obvious example is that of a Negro who has a high need value for social recognition but a low expectancy of being pledged by a fraternity.

"Greeks" vs. "Independents." Observation of fraternity and sorority life led to the hypothesis that the mean social recognition score of fraternity and sorority affiliated students will be significantly higher than the mean social recognition score of independent students. Validation groups consisted of 55 fraternity and 119 sorority members who were compared with 84 male and 88 female independents respectively.

All subjects were enrolled in two introductory classes and are comparable in terms of age, marital status (all single), and class standing. Critical ratios, means, and sigmas are shown in Table 7.

In the case of both men and women the hypothesis is supported, i.e., significant differences occur in the predicted direction.

Sex differences. The normative findings involving sex differences shown in Table 1 constitute part of the validity data for the test. In each case, the direction of the difference makes psychological

TABLE 7
DIFFERENCES IN MEAN SOCIAL RECOGNITION
SCORES BETWEEN "GREEK" AND
"INDEPENDENT" STUDENTS

Group	<i>M</i>	<i>SD</i>	<i>N</i>	<i>CR</i>
Sorority Women	29.50	8.89	119	
Independent Women	26.35	8.68	88	2.54**
Fraternity Men	31.25	9.29	55	
Independent Men	25.43	7.92	84	3.80*

* Significant at greater than the .01 level.

** Significant between the .01 and .05 level.

cal sense. On a cultural basis it seems apparent that college women in general would tend to have higher need values than men for social love and affection goals. It could also be argued that as a result of cultural training the goals associated with academic recognition would have greater reinforcement values for men than for women. The direction of sex differences is construed as supporting the construct validity of the test since it is consistent with the general cultural expectations.

Grades. Another validity measure, which can be considered as being within the known group area, was obtained by correlating grade point averages with academic recognition scores over 77 males and 64 females. The Pearson product moment correlation coefficients, the

means, and sigmas are presented in Table 8.

The low positive relationship found for males supports the notion that all other things being equal, people with higher academic recognition need values will make higher grades. The failure of the females to support the hypothesis suggests a limitation in our understanding of the construct need value. As also indicated by the sex differences in the normative data, women as a group tend to respond differently to this test than do men. The problem of sex differences was not, and should have been, taken into account in the construction of the test and in selecting methods to evaluate the adequacy of the inventory. Consideration of sex differences would seem required of both theoretical formulations and techniques of measurement involving culturally derived personality constructs.

Relationships with Edwards' Personal Preference Schedule (PPS)

The final approach to validity reported in this chapter involves the investigation of the relationships between certain scores on our scale with specified PPS (7) scores. Edwards' PPS is a relatively reliable forced-choice inventory purporting to measure 15 independent personality variables. A major attribute of Edwards' test is the rigorous control of

TABLE 8
PEARSON PRODUCT MOMENT *r*'s BETWEEN ACADEMIC RECOGNITION SCORES
AND GRADE POINT AVERAGES

Group	Grade Point		Academic Recognition			<i>r</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>N</i>	
Women	2.16	.64	28.39	7.22	64	.00
Men	2.00	.74	32.05	8.12	77	.35**

** Significant at the .05 level.

TABLE 9
PEARSON PRODUCT MOMENT r 's BETWEEN RECOGNITION (GPI) AND ACHIEVEMENT (PPS)

Group	GPI		PPS		N	r
	Recognition		Achievement			
	M	SD	M	SD		
Women	36.53	10.10	12.84	3.93	68	.42*
Men	38.00	10.30	16.15	4.37	27	.43**

* Significant at greater than the .01 level.

** Significant between the .01 and .05 level.

the social desirability factor.

The PPS provides measures on two variables, achievement and nurturance, whose definitions make them comparable to the recognition-status and love and affection variables respectively as defined by Rotter.

The similarity between these definitions leads to a predicted correlation between recognition-status as measured by the test of goal preferences and achievement as measured by the PPS, depending on the adequacy of PPS items in relation to the definition. By the same token, a correlation is predicted between love and affection and nurturance as measured by these two scales respectively.

Ninety-five students had taken the PPS in connection with another study some weeks before the GPI was administered. As shown in Tables 9 and 10 the hypothesized relationships are found for both sexes in the predicted direction.

In general, the results obtained thus far support the notion that within the tested population the GPI is a reasonably reliable and valid indicator of relative strength of need value for three of the four needs incorporated in this inventory. In analyzing the results the following suggestions were made concerning possibilities for improving this instrument: (a) elimination of the academic love and affection statements, (b) use of a more representative sample of judges to select the behavioral referents, (c) better control of social desirability, and (d) more intensive investigation of the area of sex differences.

Suggestions for Further Study

Major advances in the development of this kind of test depend upon increased understanding of the nature of the constructs being measured. The use of such an instrument in psychological research

TABLE 10
PEARSON PRODUCT MOMENT r 's BETWEEN LOVE AND AFFECTION (GPI) AND NURTURANCE (PPS)

PEARSON PRODUCT MOMENT / r S						
Group	GPI		PPS		N	r
	Love and Affection		Nurturance			
			M	SD		
	M	SD	M	SD		
Women	43.50	10.09	17.06	4.60	68	.47*
Men	42.37	10.00	14.41	5.04	27	.35**

* Significant at greater than the .01 level.

** Significant between the .01 and .05 level.

provides a means for gaining the knowledge necessary to improve the test by improving the theory underlying it. From the many research problems which could be undertaken involving the use of the Goal Preferences Inventory, a number are listed below which seem of immediate importance.

1. Inspection of the scoring profiles suggests that there are three major patterns of response to the items. First, there is a group of individuals who seem to be *need dominated*, i.e., they tend to pick all recognition items or all love and affection statements regardless of the situational context. The opposite is a much smaller group who appear to be *situation dominated*, i.e., they tend to choose all social or all academic statements regardless of the need context. The largest group appears to be composed of those who react to the items in terms of the *need-situation*. They tend to pick love and affection items in social situations and recognition items in academic situations. Factor analytic techniques which evaluate the similarity between test profiles could be employed to determine whether or not the groupings suggested above exist.

2. If groups could be differentiated in terms of the stable response patterns mentioned above, then further personality differences among these groups could be investigated. For example, it may be that the need and/or situation dominated groups are more maladjusted in college than the need-situation group. A study in progress at the University of Colorado is testing this hypothesis by investigating the relationships between scores on Rotter's (15) Incomplete Sentence Blank (a semiprojective measure of maladjustment) and scoring patterns on the GPI.

3. A prominent issue in psychology is the question of the relationships between so-called unconscious aspects of motivation and needs which subjects verbalize. A comparison of indirect measures of motivational strength, devised by McClelland (12) and his associates, with certain scores on the Goal Preference Inventory may yield some fruitful insights into this problem.

4. The assumption concerning the use of the inventory as a means of validating the theory can be tested by selecting subjects on the basis of inventory scores to participate in an experimental test of hypotheses deduced from SLT. A pilot study⁴ was conducted to test the hypothesis that if two groups are frustrated on a particular task, the group having greater need value (i.e., higher scores on the GPI) for the goals associated with success on that task would react more strongly to frustration than the group having lesser need values.

In spite of the small number of subjects and the exploratory nature of the design, this study does provide substantially more information regarding the role of need value in the reinforcement process as formulated by SLT than previous studies (2, 11) not using the GPI as

⁴ Since the primary purpose of this publication is to discuss the construction and evaluation of the GPI, only a brief abstract of one research effort done with the GPI is presented to indicate its usefulness in this respect. A full account of the study may be obtained by writing the author.

It may be worth while to emphasize that the results of this study do not validate or invalidate the GPI in the traditional sense of the term. Insofar as the test is a reliable indicator of subject variation in response to test items, it stands as a psychological measure of some variable. Research of this kind helps us better understand what and how useful that variable is. This approach, in a sense, is the essence of construct validity described by Cronbach and Meehl (4).

a selection instrument. Also important is the finding that this study indicates certain limitations in the GPI as a measure of need value, especially in the case of women. This kind of research, then, can be a basis for improving our indicators as well as testing hypotheses.

SUMMARY

The present study is based upon two related assumptions: (a) that the use of a formal behavior theory can be a powerful aid in developing a personality inventory, and (b) that a test developed in this manner can aid in bringing about a more valid and powerful conceptualization of the theory.

Demonstration of the fruitfulness of the behavior theory approach to test construction requires the building of a valid and useful measuring instrument based upon a formal psychological model. Rotter's Social Learning Theory (SLT) provided the constructs to be measured, the content and form of the items, and indicated some of the psychometric methods.

The personality inventory resulting from this study is built on the forced-choice procedure, i.e., each item consists of a pair of behavioral referents from which the subject chooses one. The two referents which are paired in any item always represent two different need categories, so each item enables a choice of need referents by a subject. This is the paradigm for the test and the model for measuring need value when expectancy is controlled. The only control of expectancy in this inventory is by way of instructions.

The inventory (named the Goal Preference Inventory, GPI) is designed to measure the relative need value of recognition vs. love and affection. It does this

by employing four subneeds: Recognition in Academic Situations (AcR), Recognition in Social Situations (SoR), Love and Affection in Academic Situations (ALA), Love and Affection in Social Situations (SLA).

Judges familiar with the college culture were asked to list behaviors they thought undergraduate students use to satisfy the four need categories. Those behaviors having high interjudge agreement were used to make up the test items. There are 20 different referents for each subneed. When the four subneeds are paired against each other (the 20 referents of one subneed against the 20 referents of another), they yield six subscales of the test, each subscale containing 20 items.

An item analysis was used to produce fairly homogeneous subtests and only those items which appeared relatively immune to being answered on socially desirable grounds were retained. The normative sample consists of 233 women and 157 men, all unmarried, undergraduate students at the University of Colorado. Split-half and test-retest methods indicate that all scales with the exception of subtest VI (ALA vs. SLA) demonstrate sufficient internal consistency and stability to warrant their use as indicators of specific psychological variables, presumably strength of need values.

A factor analysis shows three relatively independent factors corresponding to three need categories. The nonemergence of an academic love and affection factor taken in conjunction with other evidence suggests that the need for academic love and affection as defined by the GPI is not appropriate for the tested population.

Known group methods and correla-

tions with scores on Edwards' Personal Preference Schedule were used to investigate the relationships between the constructs of the inventory and external variables which should hypothetically be related to the inventory variables in specified ways. In general, the results support the notion that within the tested population the GPI is a reliable and

valid indicator of strength of need value for three of the four needs incorporated in the inventory.

On the basis of the results it seems plausible to conclude that the GPI can be useful in psychological research. The limits of its usefulness will be determined by the diversity of experimental studies employing it.

REFERENCES

1. BUTLER, J. M. The use of a psychological model in personality testing. *Educ. psychol. Measmt*, 1954, 14, 77-89.
2. CHANCE, JUNE. Generalization of expectancies as a function of need-relatedness. Unpublished doctoral dissertation, Ohio State Univer., 1952.
3. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, 10, 3-31.
4. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, 52, 281-302.
5. EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, 37, 90-95.
6. EDWARDS, A. L., & THURSTONE, L. L. An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 1953, 17, 169-180.
7. EDWARDS, A. L. *Edwards personal preference schedule manual*. New York: The Psychological Corp., 1954.
8. FLANAGAN, J. C. The use of comprehensive rationales in test development. *Educ. psychol. Measmt*, 1951, 11, 151-155.
9. GORDON, L. V. Validities of the forced choice and questionnaire methods of personality measurement. *J. appl. Psychol.*, 1951, 35, 407-412.
10. GORDON, L. V. Some interrelationships among personality item characteristics. *Educ. psychol. Measmt*, 1953, 13, 264-272.
11. JESSOR, R. A methodological investigation of the strength and generalization of verbal reinforcement. Unpublished doctoral dissertation, Ohio State Univer., 1952.
12. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton Century-Crofts, 1953.
13. PEAK, HELEN. Problems of objective observation. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden Press, 1953. Pp. 243-300.
14. ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
15. ROTTER, J. B., & RAFFERTY, JANET. *The Rotter incomplete sentences blank*. New York: The Psychological Corp., 1950.
16. THURSTONE, L. L. *Multiple factor analysis*. Chicago: Univer. of Chicago Press, 1947.
17. TRAVERS, R. M. W. Rational hypotheses in the construction of tests. *Educ. psychol. Measmt*, 1951, 11, 128-137.

(Accepted for publication January 15, 1957)

APPENDIX

THE GOAL PREFERENCE INVENTORY

This is a test to find out which kinds of activities different people prefer in certain situations while attending college. Each item consists of a pair of alternatives lettered *a* or *b*. Please choose the one activity of each pair (*and only one*) which you most prefer to do, if there is an *equal* opportunity to do either one. Be sure to select the one you actually *would* choose rather than the one you think you should choose if you had to make a choice. This is a measure of preference; there are no right or wrong answers.

DIRECTIONS

Your answers to the items on this test are to be recorded on a separate answer sheet which is loosely inserted in the booklet. **REMOVE THIS ANSWER SHEET NOW.** Print your name on the answer sheet, then finish reading these directions. Do not open the booklet until you are told to do so.

Please answer these items *carefully* but do not spend too much time on any one item. Be sure to find an answer for every choice. Find the number of the item on the answer sheet and black in the space under the letter which you choose as the activity most preferred.

Many times you will find that the *same* activities are paired in *different* combinations throughout the test.

Example: One item may read: I most prefer to:

- a) attempt to persuade a group to my political point of view.
- b) withdraw from the argument maintaining my own private point of view.

Whereas another item may read: I most prefer to:

- a) attempt to persuade a group to my political point of view.
- b) take a group of children to a movie.

Be sure to judge every item independently when making your choice. Avoid the tendency to automatically repeat the previous choice on a new item.

In some instances you also may discover that you very much like to do both things or prefer to do neither. In such cases, be sure to select the *one* you'd most rather do; the activity which is most important to you.

ANSWER SHEET*

Scoring Key: Alternative marked X refers to behavior referent for first-named need category of each subtest.

Item No.	a	b	Item No.	a	b	Item No.	a	b	Item No.	a	b	Item No.	a	b	Item No.	a	b
1.		x	21.		x	41.		x	61.		x	81.		x	101.		x
2.	x		22.		x	42.	x		62.		x	82.		x	102.	x	
3.	x		23.	x		43.		x	63.	x		83.	x		103.		x
4.	x		24.		x	44.	x		64.	x		84.		x	104.	x	
5.		x	25.	x		45.	x		65.	x		85.		x	105.		x
6.	x		26.	x		46.	x		66.		x	86.	x		106.	x	
7.		x	27.		x	47.	x		67.		x	87.	x		107.	x	
8.	x		28.	x		48.	x		68.	x		88.	x		108.		x
9.	x		29.	x		49.	x		69.		x	89.	x		109.	x	
10.	x		30.		x	50.	x		70.		x	90.		x	110.	x	
11.	x		31.	x		51.		x	71.	x		91.	x		111.	x	
12.	x		32.	x		52.	x		72.	x		92.		x	112.		x
13.	x		33.		x	53.		x	73.	x		93.		x	113.		x
14.	x		34.	x		54.	x		74.	x		94.		x	114.	x	
15.		x	35.	x		55.		x	75.	x		95.		x	115.		x
16.	x		36.		x	56.		x	76.		x	96.		x	116.	x	
17.		x	37.		x	57.	x		77.		x	97.	x		117.		x
18.		x	38.		x	58.	x		78.	x		98.	x		118.		x
19.		x	39.		x	59.	x		79.		x	99.		x	119.	x	
20.	x		40.		x	60.		x	80.	x		100.	x		120.	x	

* For testing purposes standard IBM answer sheets were used. The items are grouped in blocks of six and each block has one item from each subtest arranged in consecutive order.

REMEMBER:

Select the alternative which you most prefer to do, if there is an *equal* opportunity to do either one.

I most prefer to:

1. a) try to get into plays, on radio, etc.
b) volunteer for oral reports in the classroom.
2. a) be praised by the teacher for my ability in class.
b) be praised by the students for assisting them with academic problems.
3. a) take a course where the instructor teaches on a very high level.
b) be patient with and tolerant of a friend's more objectionable characteristics.
4. a) try to get into plays, on radio, etc.
b) sympathize with those who receive low grades.
5. a) go out of my way to help others.
b) work to acquire the social graces, i.e., activities approved by members of my group.
6. a) study for a test with others.
b) do things with the gang just because I like them.
7. a) be an expert on the best restaurants, night clubs, etc.
b) be an expert in an academic argument.
8. a) study hard in order to get good grades.
b) help the academically poorer students.
9. a) be named the senior most likely to succeed.
b) enjoy my friends' social accomplishments.
10. a) try to be seen frequently with impressive dates.
b) study for a test with others.
11. a) drive a high-priced, good-looking car.
b) volunteer my car to take people places even if it is out of my way.
12. a) support statements made in class discussions by other students.
b) suppress activities and thoughts which I think may offend others.
13. a) make an effort to rationalize or avoid discussing a poor grade.
b) give the impression that I have important connections.
14. a) know the academic standing of various departments and universities.
b) minimize the differences in ability between myself and my classmates.
15. a) spend time with other people even to the point of inefficient use of time.
b) make an effort to rationalize or avoid discussing a poor grade.
16. a) learn as many different dance steps as possible.

- b) lend my notes to a friend to study for a test we're both taking.
17. a) take the blame for difficulties when it will help a friend.
b) speak up as the voice of the group.
18. a) avoid expression of hostility in face-to-face situations.
b) avoid discussing my good grades with other students.
19. a) drive a high-priced, good-looking car.
b) be named the senior most likely to succeed.
20. a) sit near the front of the room.
b) sit with friends in class.
21. a) suppress activities and thoughts which I think may offend others.
b) compare my grades with the grades of other students.
22. a) give other students credit in class for the ideas which they have expressed.
b) speak up as the voice of the group.
23. a) work for leadership roles.
b) offer frequently to do things for people.
24. a) be patient with and tolerant of a friend's more objectionable characteristics.
b) take a course with my best friend.
25. a) work hard on weekends and holidays.
b) play the life of the party in social situations.
26. a) work to achieve membership in honorary academic societies.
b) study for a test with others.
27. a) avoid making myself stand out in social groups.
b) volunteer for oral reports in the classroom.
28. a) give the impression that I have important connections.
b) sit with friends in class.
29. a) discuss and evaluate the social competence of others.
b) avoid expression of hostility in face-to-face situations.
30. a) laugh frequently at others' jokes.
b) play down my own high grades.
31. a) use an extensive and complex vocabulary.
b) keep up with the latest slang.
32. a) work hard on weekends and holidays.
b) work with others on class projects.
33. a) volunteer my car to take people places even if it is out of my way.
b) bring in outside material to show to instructors.
34. a) accept a date for prestige—little personal interest.

- b) agree with classmates' opinions of course and instructor.
35. a) get involved in campus political activities.
b) avoid making myself stand out in social groups.
36. a) enjoy my friends' social accomplishments.
b) sympathize with those who receive low grades.
37. a) try to be seen frequently with impressive dates.
b) prepare assignments with a personal interest or involvement.
38. a) praise those who receive high grades.
b) compare my grades with the grades of other students.
39. a) lend personal belongings of substantial value.
b) read more on a subject than is required by the instructor.
40. a) be praised by the students for assisting them with academic problems.
b) be a leader in my group.
41. a) avoid pressing people for money they owe me.
b) go to important social functions which have prestige value.
42. a) help the academically poorer students.
b) offer frequently to do things for people.
43. a) work to acquire the social graces, i.e., activities approved by members of my group.
b) read more on a subject than is required by the instructor.
44. a) discuss my test results personally with an instructor.
b) avoid discussing my good grades with other students.
45. a) Be an expert in academic arguments.
b) compliment others on their dress and grooming.
46. a) Work for leadership roles.
b) minimize the differences in ability between myself and my classmates.
47. a) learn as many different dance steps as possible.
b) enjoy my friends' social accomplishments.
48. a) be named the senior best liked by my classmates.
b) go out of my way to help others.
49. a) argue with the instructor about test items which I have missed.
b) get involved in campus political activities.
50. a) start conversations about academic subjects outside of class.
b) exchange comments with nearby students during class.
51. a) offer frequently to do things for people.
b) prepare assignments with a personal interest or involvement.
52. a) play the life of the party in social situations.
b) avoid making myself stand out in academic situations.
53. a) laugh frequently at others' jokes.
b) know the social ranking of social organizations on campus.
54. a) sit with friends in class.
b) help get dates for my friends.
55. a) be well dressed on campus.
b) study hard in order to get good grades.
56. a) make a conscious attempt not to appear too bright.
b) bring in outside material to show to instructors.
57. a) know the academic standing of various departments and universities.
b) openly express my appreciation for others.
58. a) drive a high-priced, good-looking car.
b) be named the senior best liked by my classmates.
59. a) try to be seen frequently with impressive dates.
b) do things with the gang just because I like them.
60. a) listen sympathetically to a friend's problems.
b) work with others on class projects.
61. a) talk about my dates in dorms, etc.
b) tell an instructor about related work I'm doing either in other courses or on my own.
62. a) give other students credit in class for the ideas which they have expressed.
b) use an extensive and complex vocabulary.
63. a) work hard on weekends and holidays.
b) avoid pressing people for money they owe me.
64. a) get involved in campus political activities.
b) help the academically poorer students.
65. a) accept a date for prestige—little personal interest.
b) lend personal belongings of substantial value.
66. a) change my opinions and behaviors to conform with those of friends.
b) agree with classmates' opinions of course and instructor.
67. a) speak up as the voice of the group.
b) compare my grades with the grades of other students.
68. a) volunteer for oral reports in the classroom.
b) avoid making myself stand out in academic situations.

69. a) take the blame for difficulties when it will help a friend.
b) sit near the front of the room.
70. a) support statements made in class discussions by other students.
b) work to acquire the social graces, i.e., activities approved by members of my group.
71. a) try to get into plays, on radio, etc.
b) listen sympathetically to a friend's problems.
72. a) praise those who receive high grades.
b) compliment others on their dress and grooming.
73. a) know the academic standing of various departments and universities.
b) know the social ranking of social organizations on campus.
74. a) prepare assignments with a personal interest or involvement.
b) support statements made in class discussions by other students.
75. a) work to achieve membership in honorary academic societies.
b) do things with the gang because I like them.
76. a) make a conscious attempt not to appear too bright.
b) discuss and evaluate the social competence of others.
77. a) change my opinions and behaviors to conform with those of friends.
b) give the impression that I have important connections.
78. a) give other students credit in class for the ideas which they have expressed.
b) volunteer my car to take people places even if it is out of my way.
79. a) accept a date for prestige—little personal interest.
b) sit near the front of the room.
80. a) make an effort to rationalize or avoid discussing a poor grade.
b) play down my own high grades.
81. a) listen sympathetically to a friend's problems.
b) discuss my test results personally with an instructor.
82. a) take a course with my best friend.
b) know the social ranking of social organizations on campus.
83. a) be a leader in my group.
b) suppress activities and thoughts which I think may offend others.
84. a) avoid making myself stand out in social groups.
b) avoid making myself stand out in academic situations.
85. a) work for leadership roles.
b) work to achieve membership in honorary academic societies.
86. a) take a course where the instructor teaches on a very high level.
b) take a course with my best friend.
87. a) start conversations about academic subjects outside of class.
b) avoid turning down a proposed social invitation.
88. a) keep up with the latest slang.
b) avoid discussing my good grades with other students.
89. a) attempt to get to know the student leaders on campus.
b) be patient with and tolerant of a friend's more objectionable characteristics.
90. a) spend time with other people even to the point of inefficient use of time.
b) exchange comments with nearby students during class.
91. a) be praised by the instructor for my ability in class.
b) be a leader in my group.
92. a) be named the senior best liked by my classmates.
b) be named the senior most likely to succeed.
93. a) go out of my way to help others.
b) study hard in order to get good grades.
94. a) play down my own high grades.
b) attempt to get to know the student leaders on campus.
95. a) compliment others on their dress and grooming.
b) talk about my dates in dorms, etc.
96. a) take the blame for difficulties when it will help a friend.
b) minimize the differences in ability between myself and my classmates.
97. a) bring in outside material to show to instructors.
b) learn as many different dance steps as possible.
98. a) argue with the instructor about test items which I have missed.
b) avoid raising my hand every time I know the answer to a question.
99. a) laugh frequently at others' jokes.
b) use an extensive and complex vocabulary.
100. a) be well dressed on campus.
b) praise those who receive high grades.
101. a) help my friends get dates.
b) be an expert on the best restaurants, night clubs, etc.
102. a) avoid raising my hand every time I know the answer to a question.
b) openly express my appreciation for others.

103. a) discuss and evaluate the social competence of others.
b) start conversations about academic subjects outside of class.
104. a) read more on a subject than is required by the instructor.
b) sympathize with those who receive low grades.
105. a) avoid expression of hostility in face-to-face situations.
b) argue with the instructor about test items which I have missed.
106. a) go to important social functions which have prestige value.
b) work with others on class projects.
107. a) keep up with the latest slang.
b) openly express my appreciation for others.
108. a) lend personal belongings of substantial value.
b) lend my notes to a friend to study for a test we're both taking.
109. a) discuss my test results personally with an instructor.
b) attempt to get to know the student leaders on campus.
110. a) be an expert in an academic argument.
b) agree with classmates' opinions about course and instructor.
111. a) tell the instructor about related work I'm doing either in other courses or on my own.
b) change opinions and behaviors to conform with those of friends.
112. a) exchange comments with nearby students during class.
b) talk about my dates in dorms, etc.
113. a) avoid turning down a proposed social invitation.
b) be well dressed on campus.
114. a) make a conscious attempt not to appear too bright.
b) avoid pressing people for money they owe me.
115. a) go to important social functions which have prestige value.
b) take a course where the instructor teaches on a very high level.
116. a) tell the instructor about related work I'm doing either in other courses or on my own.
b) lend my notes to a friend to study for a test we're both taking.
117. a) help get dates for my friends.
b) be praised by the instructor for my ability in class.
118. a) avoid raising my hand every time I know the answer to a question.
b) be an expert on the best restaurants, night clubs, etc.
119. a) play the life of the party in social situations.
b) spend time with other people even to the point of inefficient use of time.
120. a) be praised by students for assisting them with academic problems.
b) avoid turning down a proposed social invitation.

Personality Outcomes of Client-Centered
Child Therapy¹

ELAINE DORFMAN

Children's Hospital of Philadelphia

WHILE the practice of child therapy from the client-centered viewpoint has had a parallel development with that of adult counseling, research in child therapy lags far behind. Thus far, there has been no objective study of personality outcomes of individual client-centered child therapy. Such a study would need to employ varied measures of outcome, since no single perfect criterion exists. It would also have to deal with the problem of adequate experimental controls. The project reported here had as its aim the assessment of therapy outcomes by personality tests, therapist judgments, and client follow-up statements.

HYPOTHESIS

The central hypothesis is that personality changes occur during a therapy

¹ This is one of a series of studies conducted at the University of Chicago Counseling Center, with the financial support of the Medical Sciences Division of the Rockefeller Foundation. Grateful acknowledgment of this support is made.

This report is based upon a doctoral dissertation submitted to the Committee on Human Development of the University of Chicago. The early stages of the research were guided by a thesis committee consisting of Carl R. Rogers (Chairman), Julius Seeman, and Morris I. Stein. The final committee consisted of Carl R. Rogers, William F. Soskin, and John M. Shlien.

Four persons generously gave time to judging parts of the data for reliability studies: Natalie R. Haimowitz, Wayne J. Oulton, Esselyn C. Rudikoff, and Rolland Tougas. The writer is also grateful to Mary D. Mulroy, Principal in the Chicago public school system, for many kindnesses.

period which do not occur in the same child during a no-therapy period, and which do not occur in control cases. In statistical terms, we test the null hypothesis that there are no differences significant at the 5% level between therapy and nontherapy changes. While this is the only formally stated hypothesis, the study's design implies two subsidiary ones: (a) that therapy can be conducted by an outsider in a school setting; (b) that child therapy is possible without parent therapy.

EXPERIMENTAL DESIGN

The basic design is of the pretest and posttest variety. Two types of experimental control are used. First, essential personality variables are controlled by studying the experimental group over a 13-week interval prior to therapy. This enables comparison of test score changes during a no-therapy and a therapy period for each individual child, in an experiment in perfectly matched pairs. Because random assignment of cases to separate control and experimental groups is not feasible when the number of therapy candidates is small, and because the relevant variables for personality matching are unknown, the own-control method is employed. This has previously been done in studies by Bills (4) on the effects of play therapy on reading retardation, and in research on client-centered adult counseling reported under the editorship of Rogers and Dymond (21).

A second type of experimental control is that for time, and for this a separate group of subjects is required. Because the length of therapy is not predictable, there exists no way, short of the clinically undesirable one of interruption of the therapy period, to equate the length of pretherapy and therapy intervals. Further, our pretherapy interval is a summer vacation period,

whereas therapy was conducted during the school year. Study of control subjects during an in-school period is thus required.

Our attempt to provide a control group interval equal in length to the therapy period has been imperfectly realized. Because all test results were to be blindly evaluated at the study's close, prior matching of control and therapy cases was not possible. Therefore, the average length of therapy had to be estimated in planning the time of posttesting the control cases. The estimate was 23 weeks, which proved to be 5.5 weeks too short. However, the time involved is closer to the therapy period's 28.5 weeks than is the 13-week pretherapy period.

We tried to match the control and experimental groups for initial test status (pretherapy rather than pre-wait in the case of the experimental group). Since more than one test is involved, plus the factors of age and sex, it was not possible to use precisely the same control group for each instrument. The control subjects thus vary somewhat with the test considered, although some overlapping exists.

We might, of course, have used instead of test scores some criterion such as intelligence or socioeconomic status to match experimentals and controls. This would have permitted the use of a single control group throughout. But matching requires that the correlation between matching and dependent variables be high enough to offset the loss in degrees of freedom for subsequent data analysis (10, pp. 288-295). We judged that test scores were more likely to provide such a correlated matching variable, thus increasing the efficiency of the experiment.

The research design also includes a follow-up study of 15 experimental cases, re-evaluated a year after the therapist had left the school. The follow-up interval varies from one year to one and one-half years, depending upon when each case ended therapy.

SUBJECTS

All therapy and control cases are from one public elementary school located in a prosperous middle-class neighborhood of Chicago.

Experimental Cases

The therapy group consists of 17 children considered at least normally intelligent, 12 boys and 5 girls whose teachers believed them to be maladjusted. Eighteen children began therapy, but 1 was dropped from the study owing to strong

presumption of mental defect, and because an error of scheduling assigned him to special tutoring and therapy at the same time, so that he missed many sessions.

Four criteria determined inclusion in the therapy group: (a) the teacher's judgment that the child was among the five most maladjusted in his class; (b) the school principal's concurrence that he was emotionally disturbed; (c) the consent of at least one parent; (d) age limits of 9 to 12 years. The last ensured that all tests would be appropriate for all cases. Test scores were not used in selection.

Control Cases

The controls are not therapy candidates, but are matched for age, sex, and test scores. For the Rogers Test, matching of individuals was not possible, but the groups are equated for means and variances on Total Scores. On the Sentence Completion Test, individuals are matched for their Mean Adjustment Rating (a mean of 10 rated characteristics).

INSTRUMENTS

Three personality tests were given: one objective, one nonverbal projective, and one verbal projective. Fifteen therapy children also wrote follow-up letters to the therapist.

Rogers Test of Personality Adjustment

This is a standardized objective-type paper and pencil test (20) yielding five indices of maladjustment: Total Score, Personal Inferiority, Social Maladjustment, Family Maladjustment, and Daydreaming.

Machover Human Figure Drawing Test

This is a nonverbal projective method (16), in which drawings of male and female figures are interpreted in psychoanalytic terms. While given to all subjects, this test remains un-

analyzed, owing to our failure to develop an objective scoring system.

Sentence Completion Test

A 30-item test was assembled for this study, based in part upon versions published by Rohde (22), Rotter and Willerman (23), and Shor (26). Our detailed scoring manual (7, pp. 120-188) is not included in this report, but a summary of the scoring method and the results of reliability studies are given in the Appendix.

The Sentence Completion Test provides two kinds of data: (a) a theme and attitude analysis based upon the work of Pugh (18, pp. 203-231); (b) an adjustment rating scale for 10 personality characteristics, adapted from Goldberg (12, pp. 11-51) and Reader (19, pp. 94-132). The Mean Adjustment Rating, an average of ratings on the 10 characteristics, is our index of general adjustment status.

Follow-Up Letters

This fourth instrument was adapted from Axline (2). Fifteen children of the 17 therapy cases were asked to write to the therapist telling what they remembered of the sessions and how they were currently getting along. The manual of directions for analyzing the letters (7, pp. 189-192) is not included in this monograph, but a summary of the method and reports of reliability studies are given in the Appendix.

The letters are rated along the two dimensions of memories of the therapy sessions and reports of life status. Refer to the Appendix for definitions of these terms.

EXPERIMENTAL PROCEDURES

Therapy

All 17 cases were seen individually at school by the writer, in a large basement room formerly used for classes. Therapy was conducted within the framework of the client-centered approach, following principles outlined by Axline (1) and Dorfman (6). Sessions were held weekly, except for one child who came twice weekly at his own request. Three children who asked to bring friends to one or more sessions were permitted to do so. The children were told that they must come 10 times, after which they could stop if they wished. One school year was the maximum allowed. As each case closed, the therapist rated it as relatively successful or relatively unsuccessful, without knowledge of test results.

Testing

No tests were given by the therapist. For the experimental group, pre-wait and pretherapy

projective tests were given in class by teachers. The Rogers Test was given by a graduate psychology student to small groups partly composed of classmates considered well-adjusted by their teachers. In this way, we tried to avoid singling out the therapy children for special attention other than therapy.

Posttherapy tests were given individually by the same psychology student who had administered the pre-wait and pretherapy Rogers Tests. The interval between termination of therapy and posttesting was never more than a week. The change from group to individual testing at this point introduces an unavoidable nonexperimental variable.

Follow-up tests were given in small groups to therapy cases and others by a new psychometrist, the previous one no longer being available. Two of the 17 therapy children were no longer enrolled in the school and were therefore not followed up.

For the control cases, projective tests were given in class by teachers, and the Rogers Test was given in small groups by the same psychometrist who tested the experimental group.

Follow-Up Letters

After the follow-up tests, each therapy child was sent a letter at school, asking him to write to the therapist, along the lines outlined in the preceding section on instruments. The letters were in individually addressed sealed envelopes. All 15 children who were still enrolled in the school answered.

Data Analysis

Since the investigator-therapist is also principal data analyst, care was required to rule out bias in projective test analysis. The first problem was formulation of a scoring method for the Sentence Completion Test. To be most useful, the method should be developed upon a group similar in important respects to that to which it would later be applied. Schoolmates of the research group were thus chosen.

After the first testing of therapy candidates and classmates, records of the former were set aside until over two years later. From the remaining 166 records, a random sample of 50 was drawn, and identifying data removed. The scoring manual is based upon study of these 50 cases. Each completion was typed on an index card, and the 1,500 cards sorted in various ways as different classificatory schemes were tried. Thus, the range of responses to any item could be studied in 50 cases, so that a fair notion of individual differences guided the writing of the scoring manual.

A year after the last therapy case ended, fol-

TABLE 1
ROGERS TEST MEAN SCORES AT EACH TESTING^a

Testing ^b	Rogers Scales				
	Total	Personal	Social	Family	Daydreaming
Experimental Group					
Pre-Wait	41.76	12.54	15.44	10.31	3.47
Pretherapy	41.46	11.99	17.18	9.29	3.00
Posttherapy	35.58	9.99	14.71	8.53	2.35
Follow-Up	35.99	9.09	15.95	7.97	2.98
Control Group					
Pretest	39.27	14.33	14.41	8.41	2.12
End-Test	37.00	12.32	14.88	7.44	2.36

^a Higher scores mean greater maladjustment.

^b $N=17$ for all testing except follow-up, where $N=15$.

low-up tests were obtained, and at this point analysis of experimental records began. All experimental and control group protocols were assigned code numbers from a table of random numbers, and names and dates removed. This allowed blind analysis, without knowledge of whether a given record was from an experimental or control case, or when the test had been given.

RESULTS OTHER THAN TEST OUTCOMES

Despite the chance to stop after 10 compulsory sessions, 10 cases were still in therapy when arbitrarily closed by the end of the school year. Posttherapy tests on these cases might also be viewed as

"in-process" measures, since therapy would probably have continued had the therapist remained available.

The average number of sessions is 19, with a range of 11 to 33 contacts. These occurred over a period averaging 28.5 weeks, with a range of 12 to 34 weeks. The t test shows no sex difference in therapy length.

Ten cases were rated by the therapist as relatively successful and seven as relatively unsuccessful. The judgment of "relatively successful" means only that some degree of movement was noted,

TABLE 2
ROGERS TEST MEAN SCORE CHANGES WITHIN EACH PERIOD^a

Period ^b	Rogers Scales				
	Total	Personal	Social	Family	Daydreaming
Experimental Group					
Pretherapy	-0.30	-0.55	1.74	-1.02	-0.47
Therapy	-5.88	-2.00	-2.47	-0.76	-0.65
Follow-Up	0.16	-0.54	0.75	-0.50	0.45
Control Group					
Control	-2.27	-2.01	0.47	-0.97	0.24

^a Negatively signed score changes indicate improved adjustment.

^b $N=17$ for all periods except follow-up, where $N=15$.

TABLE 3

ROGERS TEST MEAN SCORES^a AT EACH TESTING OF SUCCESSFUL AND UNSUCCESSFUL SUBGROUPS

Testing	Rogers Scales				
	Total	Personal	Social	Family	Daydreaming
Successful Cases ^b					
Pre-Wait	42.19	12.72	14.75	11.92	2.80
Pretherapy	43.29	11.89	18.60	9.50	3.30
Posttherapy	34.34	7.89	15.30	9.15	2.00
Follow-Up	39.11	8.57	18.46	8.80	3.28
Unsuccessful Cases ^c					
Pre-Wait	41.14	12.28	16.43	8.00	4.43
Pretherapy	38.86	12.14	15.14	9.00	2.57
Posttherapy	37.36	13.00	13.86	7.64	2.86
Follow-Up	29.76	10.12	10.94	6.30	2.40

^a Higher scores mean greater maladjustment.^b $N=10$ for all testings of successful cases.^c $N=7$ for all testings of unsuccessful cases, except for follow-up, where $N=5$.

the nature of which varied from case to case. Chi square shows no significant association between sex and therapist judgment, nor between method of termination (by therapist or child) and therapist judgment. Neither is there any association between sex and method of termination. Further, the t test shows no tendency for therapist judgment to favor either longer or shorter cases.

RESULTS OF THE ROGERS TEST OF PERSONALITY ADJUSTMENT

Table 1 gives mean scores at each testing for all Rogers scales. Higher scores mean greater maladjustment. Table 2 shows the score changes within each period of the study. In Tables 3 and 4, the results are separately presented for subgroups judged by the therapist to be relatively successful and unsuccessful.

TABLE 4

ROGERS TEST MEAN SCORE CHANGES^a WITHIN EACH PERIOD FOR SUCCESSFUL AND UNSUCCESSFUL SUBGROUPS

Period	Rogers Scales				
	Total	Personal	Social	Family	Daydreaming
Successful Cases ^b					
Pretherapy	1.10	-0.83	3.85	-2.42	0.50
Therapy	-8.95	-4.00	-3.30	-0.35	-1.30
Follow-Up	4.77	0.68	3.16	-0.35	1.28
Unsuccessful Cases ^c					
Pretherapy	-2.28	-0.14	-1.29	1.00	-1.86
Therapy	-1.50	0.86	-1.28	-1.36	0.29
Follow-Up	-9.06	-3.00	-4.06	-0.80	-1.20

^a Negatively signed score changes mean improved adjustment.^b $N=10$ for all periods for successful cases.^c $N=7$ for all periods for unsuccessful cases, except for follow-up, where $N=5$.

Score changes are studied by the t test. We begin by considering the therapy period. Then, we compare its results with those of the pretherapy period, the separate control group, and the follow-up period.

Changes During Therapy Period

For the total group of 17 cases, two of five indices show significant improvement during therapy. These are Total Score ($p < .01$) and Social Maladjustment ($p < .05$). There is a possible trend to improvement on Personal Inferiority ($p < .10$). See Table 2 for amounts of change.

Subgroup analysis shows no significant differences between changes of boys and girls, nor between self-closed and therapist-closed cases. But cases judged successful improve significantly more than unsuccessful ones on Total Score ($p < .05$) and Personal Inferiority ($p < .05$), although their pretherapy scores do not differ. See Table 4 for amounts of change. Taken by themselves, the unsuccessful cases fail to change significantly from their own pretherapy mean on any index, but the successful cases improve significantly on Total Score ($p < .01$), Personal Inferiority ($p < .01$), and Social Maladjustment ($p < .02$). This means that the total group's therapy improvements mainly reflect changes by the cases judged successful.

Returning to the total group of cases, we find no relationship between length of therapy and amount of improvement. Whether length is expressed in number of sessions or in number of weeks, ρ varies from $-.28$ to $.39$, all insignificant.² Neither is there any reliable relation be-

tween number of contacts and degree of pretherapy maladjustment on the Rogers Test. Rho varies from $-.20$ to $.46$, none significant at the $.05$ level. But trends to significance may exist for Total Score ($p < .10$) and Daydreaming ($p < .06$). All of these ρ 's may be spuriously low, since variability in number of sessions is artificially limited by the arbitrary closing of 10 incomplete cases.

Changes During Pretherapy Period

Therapy period changes cannot be ascribed to the sessions unless it can be shown that similar changes do not occur in the group during the pretherapy or control period. None of the pretherapy changes shown in Table 2 are significant, so that the hypothesis of spontaneous improvement is untenable for this group as a whole.

Subgroup analysis shows no sex differences in amount of change, nor are there differences in the relative changes of self-closed and therapist-closed cases. The latter means that the subsequent tendency to discontinue therapy cannot be ascribed to spontaneous improvement in these cases prior to therapy.

As was found true of the therapy period, there is a difference in the relative changes of cases later judged successful or unsuccessful in therapy. Among the changes shown in Table 4, only the difference on Family Maladjustment is significant at the $.05$ level, and favors the successful cases. Reverse trends may exist for Social Maladjustment and Daydreaming, where differences significant at the $.10$ level favor the unsuccessful cases. However, the unsuccessful subgroup was initially more maladjusted on Daydreaming ($p < .05$), so that its somewhat greater improvement during the pretherapy interval may reflect regression effects in the test.

² Throughout this report, ρ and W are understood to be derived from Kendall's b formulas (14), which correct for tied ranks.

Considered by themselves, the unsuccessful cases fail to depart significantly from their own pre-wait mean on any index, although a trend to improvement is found on Daydreaming ($p < .10$). In contrast, the successful cases improve on Family Maladjustment ($p < .05$), and grow worse on Social Maladjustment ($p < .05$). Taking only differences significant at the .05 level, we note that the successful cases change on two of five Rogers Test indices during the pretherapy period, showing gain on one and decrement on one. The unsuccessful cases never change significantly. This implies that it is easier for therapy to channelize existing change into the direction of better adjustment, insofar as the Rogers Test measures it, than to institute change where it is not already occurring to a significant degree.

Comparison of Changes During Pretherapy and Therapy Periods

We saw above that for the total group, certain reliable changes occur during therapy, but not within the pretherapy interval. The question now is whether therapy score shifts are significantly greater than pretherapy ones. Among the changes shown in Table 2, that on Total Score improves significantly more ($p < .05$) during therapy than prior to it. A trend to significant difference on Social Maladjustment ($p < .10$) also favors the therapy interval.

No subgroup comparisons are made for sex or termination method, since no within-period differences exist. But having found such differences in the successful and unsuccessful cases, it is appropriate to consider them further.

The 10 cases judged successful show therapy score improvements reliably greater than pretherapy changes on Total Score ($p < .05$) and social Mal-

adjustment ($p < .01$). See Table 4 for the amounts involved. The seven unsuccessful cases show no difference between their therapy and pretherapy changes. Thus, the total group's greater improvements during therapy as compared to the pretherapy interval are largely due to gains of cases which the therapist viewed as relatively successful.

Changes in the Separate Control Group

While the total experimental group shows no reliable pretherapy gain, it might be argued that the interval is too short, and that in a vacation period children are deprived of possible gains from mere school attendance. The separate control group, studied over a 23-week school period, shows no reliable gain on any Rogers index. A trend to improvement may exist on Personal Inferiority ($p < .10$). On the whole, though, it seems fair to conclude that time and school attendance, within our given limits, do not improve test performance.

Comparison of Control Group and Therapy Changes

The experimental group during therapy improves significantly more than the separate controls on Social Maladjustment ($p < .05$). A trend to significance on Total Score ($p < .10$) also favors the experimental group. Compared with the results of the therapy versus own-control period tests, we find that it is the same two indices which favor the therapy interval, although their significance levels are reversed here. Amounts of change within each period are found in Table 2.

Comparison of Control Group and Pretherapy Changes

Here we compare test changes during two no-therapy intervals. There are no reliable differences between them. Thus,

a longer interval between testings does not result in greater change, nor does the in-school out-of-school variable produce a difference. Likewise, the data indicate that children recommended for therapy are no more apt to show test changes prior to therapy than a group of noncandidates.

Changes During Follow-Up Period

Are the therapy gains lasting, or are they transitory and dependent upon the continuance of the therapeutic relationship? We answer this by comparing the posttherapy and follow-up scores of the 15 cases still in the school at follow-up time. Posttherapy and follow-up scores of these 15 children do not differ significantly. Thus, therapy gains are maintained. This is true even though 9 of the 15 cases were arbitrarily closed. Further, compared to pretherapy scores, follow-up tests are reliably better on Total Score ($p < .02$) and Personal Inferiority ($p < .02$). A trend to significance is found on Family Maladjustment ($p < .10$), which also favors the follow-up test. This indicates absence of regression to pretherapy status upon discontinuance of the therapy sessions.

There are no sex differences in follow-up changes, which agrees with results of the pretherapy and therapy periods. But comparison of follow-up changes among self-closed and therapist-closed cases shows a difference on Social Maladjustment ($p < .05$) favoring the self-closed cases. The subgroups do not differ on posttherapy Social Maladjustment. Thus it is not until the follow-up period that a difference in changes made by self-closed and therapist-closed cases occurs. But the self-closed subgroup's apparent improvement over its own posttherapy mean is not significant for any index.

The apparent regression of the therapist-closed group also lacks reliability. There may be a trend to significant regression on Social Maladjustment ($p < .10$) in the therapist-closed cases, but whether a longer follow-up interval would strengthen or reverse this trend is problematical.

Changes of successful and unsuccessful cases differ on Total Score ($p < .02$), Personal Inferiority ($p < .05$), and Social Maladjustment ($p < .02$), and all these differences favor the unsuccessful subgroup. A trend to significance on Daydreaming ($p < .10$) also favors the unsuccessful cases. The Personal Inferiority difference may reflect regression effects in the test, since there the subgroups differ on posttherapy status. But taken by themselves, neither subgroup changes reliably from its own posttherapy mean on any index. However, while the regression of successful cases is never closer to significance than the .20 level, improvement of unsuccessful cases reaches the .10 level on Total Score, Personal Inferiority, and Social Maladjustment. Here again, the question of the possible effect of a longer follow-up period can be raised but not answered by this study.

Comparison of Follow-Up and Therapy Period Changes

It has been seen thus far that the group as a whole undergoes significant changes during the therapy interval, but not during the follow-up period. But are the between-period differences themselves reliable? The data show no significant difference between therapy and follow-up period changes for the group of 15 cases. Thus, a longer interval does not necessarily result in greater score change.

The 10 successful cases show reliable differences between their follow-up and

therapy period changes on Total Score ($p < .02$), Personal Inferiority ($p < .05$), and Social Maladjustment ($p < .05$), and these favor the therapy period. The movement is toward improvement during therapy and decrement across the follow-up. But we noted earlier that their therapy improvements are reliable and that their follow-up decrements are not. This means that for this particular subgroup, the essential changes occur during therapy and not after it.

The five unsuccessful cases followed up are quite different. Their follow-up and therapy period changes show no significant difference. Further, we noted above that their therapy changes never approach significance, while their follow-up period improvements attain the .10 level on Total Score, Personal Inferiority, and Social Maladjustment. It looks as though we may possibly have beginnings of change during the follow-up period, but whether a longer interval would confirm this is uncertain.

Comparison of Follow-Up and Pretherapy Period Changes

This comparison involves two no-therapy intervals of the experimental group. There are no significant differences between them. The longer follow-up period does not result in greater change. This answers the question posed by the relative brevity of the pretherapy period in comparison to the therapy interval. Our findings render unlikely the hypothesis that the superiority of the changes during therapy is attributable to the time disparity.

Although there are within-period differences in changes of successful and unsuccessful cases, neither subgroup shows any reliable difference between its own follow-up and pretherapy changes.

Comparison of Follow-Up and Control Group Changes

Here we compare changes during two no-therapy intervals for two different groups of subjects. There are no significant differences between changes during the follow-up period and those of the control group. Thus, as in the case of the pretherapy versus separate control group comparison, a time period about twice as long does not result in greater change on the Rogers Test.

RESULTS OF THE SENTENCE COMPLETION TEST

Themes and Attitudes (see Appendix for descriptions of these) show only random fluctuations throughout the study. They appear to be highly stable measures, affected neither by time nor by therapy. This may reflect either the use of insensitive categories or true stability of underlying function.

Since there are no changes in Themes and Attitudes, we shall confine ourselves to study of the Adjustment Ratings. As Table 9 shows these to be correlated measures, we shall describe our results in terms of the Mean Rating, a general adjustment α index. Presentation of changes by individual scales would falsely imply a broader area of personality change than is justified by our overlapping measures.

Table 5 gives group averages on Mean Rating at each testing and also changes within each period of the study, for the total experimental and control groups. In Table 6, the results are separately presented for subgroups judged by the therapist to be relatively successful and relatively unsuccessful. All differences are evaluated by the t test. As with the Rogers Test, we begin with results of the therapy period. These are then compared

TABLE 5
SENTENCES TEST MEAN ADJUSTMENT RATING AT EACH TESTING AND
CHANGES WITHIN EACH PERIOD^a

Experimental Group				Control Group	
Mean Rating ^b at Each Testing					
Pre-Wait 4.08	Pretherapy 4.54	Posttherapy 3.24	Follow-Up 3.23	Pretest 4.53	End-Test 4.45
Mean Change ^c Within Each Period					
Pretherapy 0.46		Therapy -1.30	Follow-Up -0.92	Control -0.08	

^a $N=17$ for all cases except follow-up, where $N=15$.

^b Higher ratings mean greater maladjustment.

^c Negatively signed rating changes mean improved adjustment.

with those for the pretherapy period, the separate control group, and the follow-up period.

Changes During Therapy Period

During therapy, there is a significant improvement in general adjustment within the total group of cases ($p < .001$). Subgroup analysis shows no differences in relative changes of boys and girls, self-closed and therapist-closed cases, or cases judged successful and unsuccessful. Un-

like the Rogers Test, Sentences Mean Adjustment Ratings bear no relationship to therapist judgment. The absence of subgroup differences may be a function of the relatively consistent trend toward better adjustment. Where nearly everyone improves, there is less chance for subgroup differences to emerge.

In agreement with the Rogers Test, we again find no relationship between length of therapy and amount of improvement. Whether therapy length is expressed in number of sessions or in number of weeks makes no difference, the respective ρ 's being $-.08$ and $.08$. Neither is there any relationship between number of therapy sessions and degree of pretherapy maladjustment on Mean Adjustment Rating, as seen in $\rho = -.10$. As noted for the Rogers Test, these correlations may be spuriously low on account of artificially limited variability in number of therapy sessions.

Changes During Pretherapy Period

The total group shows significant adjustment decrement in the pretherapy interval ($p < .02$), as seen in a higher Mean Rating at its close (Table 5).

Subgroup analysis indicates no reliable difference between changes of boys versus

TABLE 6
SENTENCES TEST MEAN ADJUSTMENT RATING
AT EACH TESTING AND CHANGES WITHIN
EACH PERIOD FOR SUCCESSFUL AND
UNSUCCESSFUL SUBGROUPS^a

Mean Rating ^b at Each Testing				
	Pre-Wait	Pre-therapy	Post-therapy	Follow-Up
Successful	4.11	4.76	3.25	3.20
Unsuccessful	4.04	4.21	3.21	3.30
Mean Change ^c Within Each Period				
	Pretherapy	Therapy	Follow-Up	
Successful	0.65	-1.51	-0.05	
Unsuccessful	0.17	-1.00	0.40	

^a $N=10$ for successful cases. $N=7$ for unsuccessful cases, except for follow-up, where $N=5$.

^b Higher ratings mean greater maladjustment.

^c Negatively signed rating changes mean improved adjustment.

girls, self-terminated versus therapist-terminated cases, and cases judged successful or unsuccessful.

Comparison of Changes During Pretherapy and Therapy Periods

While the therapy interval is characterized by significant improvement on Mean Adjustment Rating and the pretherapy period by significant decrement, we must ask whether a reliable between-period difference exists. Mean Rating improvement during therapy is reliably greater than pretherapy change ($p < .001$).

Subgroup analyses are not made, since there are no within-period differences.

Changes in the Separate Control Group

Although therapy period improvement clearly outweighs pretherapy change, it must be kept in mind that the pretherapy interval is shorter and is an out-of-school period. The separate control group, studied over a longer in-school period, shows no reliable change in Mean Rating. Once again, a longer in-school period does not in itself result in improved adjustment.

Comparison of Control Group and Therapy Changes

While the therapy group improves during therapy and the control group does not, we must still ask whether the therapy changes are reliably greater. Therapy improvement on Mean Rating is significantly greater than that of the group of separate control cases ($p < .001$). This agrees with the results of the therapy versus own-control comparison. It indicates that the greater improvement during therapy as compared to the pretherapy interval cannot be due to duration or to school-period.

Comparison of Control Group and Pretherapy Changes

This comparison involves two no-therapy intervals which differ in time span and in the in-school out-of-school variable. The separate control cases improve more on Mean Rating than do the experimental cases prior to therapy ($p < .01$). But we saw above (Table 5) that the controls show an insignificant improvement, while the experimentals show a significant decrement. Thus, the reliable difference between their changes arises from their movement in opposite directions. It is theoretically possible that, given a longer in-school pretherapy period, the experimental group might show improvement rather than decrement in adjustment on Mean Rating. Our data cannot answer this question, but it is not crucial, because we saw in the immediately preceding section that the therapy improvement is reliably better than that of the control group. From this we conclude that even if a hypothetical longer pretherapy period resulted in equivalent pretherapy and control group changes, these would still be less than those of the therapy period.

Changes During Follow-Up Period

While therapy Mean Rating improvement is clearly greater than pretherapy and control group gains, its durability has to be established. No reliable difference exists between posttherapy and follow-up Mean Rating of 15 cases followed up. Thus, the therapy gain is maintained after the interviews cease. Moreover, follow-up Mean Rating is significantly better than at pretherapy ($p < .001$). There is thus no regression to pretherapy status upon discontinuance of therapy.

Subgroup analysis shows no reliable contrasts between follow-up period

changes of boys versus girls, self-closed versus therapist-closed cases, nor in those of cases judged to be successful or unsuccessful.

Comparison of Follow-Up and Therapy Period Changes

Thus far, we have seen that Mean Rating improves significantly during therapy, but not across the follow-up interval. We now find a reliable between-period difference which favors the therapy period ($p < .01$) for the 15 cases followed up. This means that the basic change in these cases occurs during therapy and not after it, even though the follow-up period is longer.

Subgroup analyses are not made, since there are no within-period differences for any subgroup.

Comparison of Follow-Up and Pretherapy Period Changes

The comparison here is between changes during a short no-therapy interval just prior to therapy and a longer no-therapy interval just after therapy. For the 15 cases followed up, there is no reliable difference in Mean Rating changes between pretherapy and follow-up periods. Once again, a longer in-school period does not result in greater improvement. This is further evidence that the superiority of the therapy period gains, relative to those of the brief pretherapy period, cannot be explained by the time difference.

Subgroup analyses are not offered because there are no within-period subgroup contrasts on Mean Rating changes.

Comparison of Follow-up and Control Group Changes

Here we compare changes during two no-therapy intervals, for the experi-

mental and control groups. There is no reliable difference in Mean Rating changes between follow-up and control group, and this despite the fact that the former period covers a time interval at least twice as long.

RESULTS OF THE FOLLOW-UP LETTERS

Ten letters recall the therapy sessions with pleasure and are thus classified Positive for therapy memories; two are hostile or Negative; three are noncommittal or Neutral. A chi-square test of the hypothesis of zero difference among categories results in a value of $\chi^2 = 7.60$ ($p < .05$). Thus, the obtained distribution, in which positive memories predominate, differs significantly from chance expectancy.

Nine letters exhibit a positive sense of well-being and are therefore classified Positive for reported life status; two express unhappiness or failure and are classified Negative; four are noncommittal or Neutral. A chi-square test of the hypothesis of zero difference among categories gives $\chi^2 = 5.20$ ($p < .10$). Thus, the obtained distribution does not differ significantly from chance expectancy, although we may speak of a trend to significance. We may note that two of the four cases classified Neutral for reported life status fail to supply any information on this question. Their letters are confined to therapy memories, and our procedure in cases of failure to respond has been to classify the letter automatically as Neutral.

The agreement between therapy memories and reported life status classifications is examined by chi square. There are six agreements and nine disagreements; with three judgment categories, chance agreement is one-third. Testing the deviation from expected agreements,

TABLE 7

RHO CORRELATIONS BETWEEN SENTENCES MEAN ADJUSTMENT RATING AND ROGERS SCALES AT EACH TESTING OF EXPERIMENTAL GROUP^a

Testing	Correlations with Each Rogers Scale				
	Total	Personal	Social	Family	Daydreaming
Pre-Wait	-.14	-.23	-.26	.17	.03
Pretherapy	-.18	-.06	-.22	-.27	-.09
Posttherapy	-.26	-.31	-.38	.17	-.16
Follow-Up	-.06	-.23	-.08	.36	-.31

^a $N=17$ for all testings except follow-up, where $N=15$.

$\chi^2 = 0.30$ ($p < .70$). The retrospective view of therapy is thus independent of reported current life status.

By inspection, the letters of the following subgroups show no differences: (a) cases having above and below the average number of therapy sessions; (b) self-closed and therapist-closed cases; (c) cases judged successful and unsuccessful; (d) boys and girls. No statistical tests are made on these enumeration data because frequencies are too small for chi square; Yates' correction is inapplicable, as more than one degree of freedom is involved.

INTERRELATIONSHIPS AMONG MEASURES

The three preceding sections present test results taken one at a time. We now examine the extent of agreement or disagreement among them. Comparisons are limited to data of the experimental group. In the section on experimental design, we indicated that it was not possible to use precisely the same control cases for the Rogers and Sentence Completion tests, so

that no between-test comparisons can be made for them. In addition, only the therapy children wrote follow-up letters.

Relationship Between Scores on Rogers and Sentence Completion Tests

Rho correlations between Rogers Test scales and Sentences Mean Adjustment Rating at each of four testings are given in Table 7, which shows no significant coefficients. Contrary to expectation, 16 of the 20 ρ 's are negative. By the one-tailed Sign Test, this is significant at the .01 level. These results, of course, are no estimate of test agreement in an unselected sample. By studying children whom teachers considered maladjusted, we have, by definition, tapped one extreme of the curve. This may well have distorted any relationships between the Rogers and Sentences tests.

Comparison of Changes on Rogers and Sentence Completion Tests

Changes on each Rogers Test scale are compared with those on Sentences Mean Adjustment Rating for each period of the study. Rho coefficients are given in Table 8, which indicates lack of relationship between measures of change. Since 15 comparisons are made, the one significant negative coefficient may be a chance effect.

TABLE 8

RHO CORRELATIONS BETWEEN CHANGES ON SENTENCES MEAN ADJUSTMENT RATING AND ROGERS SCALES WITHIN EACH PERIOD OF EXPERIMENTAL GROUP^a

Period	Correlations with Each Rogers Scale				
	Total	Personal	Social	Family	Daydreaming
Pretherapy	-.26	-.48	.23	-.50 ^b	.39
Therapy	.17	.38	-.26	-.08	.09
Follow-Up	.25	-.03	-.05	.32	-.14

^a $N=17$ for all periods except follow-up, where $N=15$.

^b $p < .05$.

Comparison of Follow-Up Letters and Tests

In order to compare the follow-up letters and tests, a numerical score for the letters is required. It is based upon the judgment method outlined in the section on instruments, since the reliability of this is known. For both therapy memories and reported life status, a value of one is assigned to a letter judged Positive, a value of two to a Neutral letter, and a value of three to a Negative one. Since each letter is evaluated for two factors, the scores can vary from two to six. The 15 letters are then ranked for degree of positive affect. Obviously, there are many ties, but allowance can be made for this (14, pp. 80-89).

Since the letters constitute a very crude index of adjustment status, we compare them only with our other indices of over-all status, Rogers Total Score and Sentences Mean Adjustment Rating.

The 15 cases are ranked at follow-up for Letters, Sentences, and Rogers Test status, and the rankings simultaneously compared by the *W* statistic. $W = .47$ ($p < .20$), indicating that if each measure taps some aspect of personal adjustment, it samples a different facet of the criterion universe. This is possible when so complex a criterion as adjustment is involved.

DISCUSSION

The absence of significant relationship between the therapist's judgment of outcome and the duration of therapy is somewhat surprising, since one might expect that a client's willingness to continue therapy would influence the therapist's view of his progress. That this is not the case here may reflect idiosyncrasy of judgment, inasmuch as the work of a single therapist is involved. This might explain the discrepancy from Cartwright's study of 78 adult client-centered cases (5), where a curvilinear relationship, positive at both extremes of length, existed between length and judged success. His results are supported by those of Taylor (28), in a study of 309 analytically oriented cases. While it may be that length of therapy is a less significant dimension among children than among

adults, a child therapy study involving several therapists would be required to test this. Moreover, our arbitrary closing of 10 cases, dictated solely by convenience, automatically restricted variability in therapy length, thus obscuring whatever correlation might actually exist between length and other factors.

Absence of sex difference in judged outcome agrees with Cartwright's findings on adult cases (5). But, in contrast with our results, Dymond (8, 9) found test changes which favor adult females. Whether this reflects a difference between children and adults or a difference between measuring instruments remains unknown. However, our lack of sex difference in test outcomes also disagrees with a study of nondirective group play therapy by Fleming and Snyder (11). They found greater Rogers Test improvements among girls. In their study, both sexes were seen by one female therapist. The difference from our study, also employing a single female therapist, may mean that the therapist's sex plays a different role in group and individual therapy. The personality of the particular therapist involved may also be a factor. These are moot points at present, but it is possible to design experiments capable of answering them. It should also be kept in mind that in our study boys outnumbered girls 12 to 5, so that sex comparisons are based upon a very small group of girls. No attempt was made to include an equal number of each sex. Boys are more frequently referred, and it was desired that the experiment reflect the conditions of therapy as ordinarily practiced.

Our Rogers Test results for the therapy period contradict those of Seeman and Edwards (25) on group child therapy with maladjusted retarded readers. Their

therapy cases grew worse on Rogers Total Score, the decrement very nearly being significantly greater than that of the control group. Their study involved daily sessions with a teacher-therapist for four months, in contrast to our own conditions of weekly sessions with a non-teacher therapist over an average of 28.5 weeks. It is unlikely that the time or group therapy differential is responsible for the discrepancy between their results and ours, since Fleming and Snyder's (11) group therapy, covering an even briefer period of six weeks, did not result in Rogers Test decrement. But if the time and the group-individual factors are not per se explanatory, this does not rule out their possible interaction effects with other conditions such as reading retardation.

What sort of children do teachers refer for psychotherapy? Referral comments as well as observations of the therapist indicate that shy and "good" children as well as aggressive and troublesome ones were included. Of the 17 cases, 16 were maladjusted on some Rogers Test index at the time of referral, and 10 were on the maladjusted side of the scale on Sentences Mean Adjustment Rating. While there was considerable variation in the group's initial test performance, the teachers' opinions generally appear reasonable.

We have noted statistically significant therapy changes on both Rogers and Sentences tests, but this does not guarantee practical significance, which depends upon the size of the change in relation to the amount of effort expended in bringing it about. If some other more economical procedure than individual client-centered therapy were to result in similar changes, our data would have little practical importance. We do not have com-

parative information on other procedures, so that an attempt to assess the amount of our changes must take another form.

On the Rogers Test, the average Total Score improvement during therapy is 5.88 points, which is 14% of the mean pretherapy score of 41.46. But the *SD* of pretherapy scores is 9.72 points, so that the mean therapy change represents less than 1 *SD*. Thus, while mean therapy change is reliable, it is small in relation to variability in pretherapy scores. On the Sentence Completion Test, the average Mean Adjustment Rating improvement during therapy is 1.30 scale intervals, which is 29% of the mean pretherapy rating of 4.54. Since the *SD* of pretherapy ratings is 0.52, the therapy change represents 2.50 *SD*. Thus, therapy change is large in relation to variability in pretherapy ratings. These comparisons imply greater change on the projective than on the objective test, which runs counter to the idea that projective tests measure the "deeper" personality. We should expect more change on an objective test, which presumably taps more superficial aspects. Perhaps the distinction is less applicable to children; but, even so, this would not explain the direct reversal of the prediction.

Let us assume for the moment that the Sentence Completion Test is a valid instrument, although this is entirely unknown. At the pre-wait examination, our average therapy subject looks more maladjusted on it than on the Rogers Test. If the projective test is indeed more sensitive to the presence of psychological disturbance, then it would not be surprising if it were also more sensitive to slight changes in such disturbance. Some evidence may be found in the pretherapy period, where the group grew reliably

worse on Sentences Mean Adjustment Rating, but not on any Rogers scale. Therefore, the projective test's showing of greater change is not restricted to the period of therapy. During the follow-up period, neither test changed reliably. We therefore infer that when a change is great enough to show at all on the Rogers Test, it is apt to be more marked on the Sentence Completion Test. This inference would have to be tested on additional samples before real confidence could be placed in it, but it is interesting enough to warrant further study.

Lack of reliable test improvements prior to therapy agrees with findings by Grummon (13) in a group of 23 adult client-centered cases studied over a 60-day no-therapy period. But our own results are more extreme: the Sentences Mean Adjustment Rating shows actual adjustment decrement. Since we found no such decrement on the Rogers Test, this suggests that some peculiarity of the Sentence Completion Test, rather than of our client group, is involved. Whether the Sentence Completion Test is more sensitive to minor changes, or whether it measures some different aspect of personality adjustment from that sampled by the Rogers Test, is uncertain.

Recent work by Barron and Leary (3) contradicts both our own and Grummon's (13) results. They found reliable MMPI improvements in a group of neurotic adults waiting for therapy, and no difference between their mean changes and those in a group undergoing analytically oriented treatment. However, ethical scruples prevented random assignment to control and experimental groups, so that those expected to regress without treatment were more likely to be placed in the treatment group. Moreover, an important difference between their study and ours is that, during the pre-

therapy period, our children had no idea that they would later receive therapy. Also, they were selected by teachers, not by self-referral, so that the variable of motivation for change is not a factor in the sense in which it might operate to produce changes in a self-selected group.

The unknown validity of the Sentence Completion Test is a serious shortcoming of this study. We had originally hoped to attack this problem by having teachers select the five best-adjusted as well as the five worst-adjusted children in their classes. A very crude validity check was to have been made by comparing protocols of these two groups. But it turned out that the children considered best-adjusted were nearly all girls, while those judged worst-adjusted were mainly boys, making it fruitless to compare their records for validity purposes. It is unlikely that these judgments simply reflect female teachers' prejudices, as it is common experience that boys tend to show more frequent emotional problems than girls (16).

One of our findings is particularly at variance with expectations from research in adult counseling. We found no therapy change in relative number of positive, negative, and neutral attitudes on the Sentence Completion Test. Seeman's protocol analysis of 10 adult client-centered cases showed a clear shift from negative to positive attitudes as therapy progressed (24). Perhaps there is a real difference in this regard between the therapy of children and adults. As a matter of fact, increased expression of negative feelings characterized the four play therapy cases studied by Landisberg and Snyder (15). Our own results do not agree with theirs, but there is an age difference of about four years involved. Moreover, attitudes expressed in therapy sessions may not have the same signifi-

cance as those shown on a psychological test. But since the present study shows reliable shifts in the adjustment ratings despite absence of change in distribution of positive, negative, and neutral attitudes, it is possible that the latter categories are too crude or general to reflect test changes.

Perhaps the greatest shortcoming of the present study is the lack of behavioral criteria of changes attributable to therapy. Improvements noted on psychological tests do not necessarily mean better interpersonal relations, freer use of intellectual capacities, more mature behavior, or the achievement of other goals generally implicit in the process of therapy. To measure these requires observation of the child outside the therapy and testing situations, which was unfortunately beyond the scope of this investigation. But now that we have evidence of some test improvements, the next step might well be the conduct of studies directed toward the problem of behavior in other situations. This is particularly important in view of the report by Teuber and Powers (29). In a large-scale study of therapy outcomes among predelinquents in a social work setting, they found no difference in later delinquency of therapy and control cases, despite the optimistic views of clients and therapists.

In a child therapy study, behavioral criteria might come from teacher and parent interviews. This was impossible in our case because the single therapist was fully occupied with the therapy sessions. Another disadvantage of the use of a single therapist is that it prevents generalization from the results, which can be a function of either the particular therapeutic system or the therapist's personality. Future research in this area might better be planned as a group proj-

ect, to overcome some of these limitations.

In general, we find that psychotherapy, within the limits of our criteria, can be effective in a school setting, and without parent counseling. This is not to deny possible benefits from such counseling. But many parents cannot accept the idea of help for themselves, even while allowing their children to receive it. If the clinician's choice lies between refusing such cases altogether and trying to work with the child alone, our results suggest that the latter alternative has beneficial possibilities.

SUMMARY AND CONCLUSIONS

This project had as its aim the assessment of personality outcomes of individual client-centered child therapy by means of psychological tests, therapist judgments, and follow-up letters. The hypothesis was that personality changes occur during a therapy period which do not occur in the same child during a no-therapy period, and which do not occur in a control group.

The basic experimental design involved observation during three time periods for the therapy group of 17 cases: (a) pretherapy or control period; (b) therapy period; (c) follow-up or post-therapy period. Seventeen control cases were tested twice, over a time interval close to the average length of therapy, as a way of controlling for time. Criterion measures were therapist judgment, an objective test, a projective test, and follow-up letters. The last two were found to have adequate intrascorer and interscorer reliability, the figures for which can be examined in the Appendix. Therapy was conducted in school by a single female therapist, the investigator, who did no testing.

The average length of therapy was seen

to be 19 sessions, with 10 of the 17 cases considered relatively successful by the therapist. The therapist's judgment was unrelated to length of therapy, method of termination (by child or therapist), or sex of the client. Nor were there any sex differences in duration of therapy or in termination method.

On the Rogers Test, we noted certain therapy improvements reliably greater than those of the pretherapy period and of the separate control group. While the process of change did not continue in the follow-up period, the gains of therapy were maintained. Test results were unrelated to the sex of the child, the duration of therapy, or to termination method, but they were related to the therapist's judgment in the case of the Rogers Test.

The Sentence Completion Test showed great stability of Themes and Attitudes, regardless of experimental conditions. Mean Adjustment Rating showed gains during therapy, and these were reliably greater than those of the pretherapy interval and the separate control group. The therapy gains were maintained across the follow-up period, but no significant improvements occurred once therapy had stopped. Subgroup comparisons were minor. Results on the Sentence Completion Test were unrelated to the therapist's judgment of outcome. Likewise, they bore no relation to the duration of therapy.

Follow-up letters showed that positive (favorable) memories of the therapy experience predominated, but that no reliable trends existed for reports of present life status. Moreover, the therapy memories and reported status were unrelated. On an inspection basis, no subgroup differences appeared.

The three measures of therapy out-

comes were found to be uncorrelated. This suggests that if each examines some aspect of personal adjustment, it is likely to be a different facet of the criterion universe.

Our conclusions are as follows.

1. Reliable test improvements occur concomitantly with a series of therapy sessions.

2. Time alone does not produce reliable improvements on our tests. While individuals may show "spontaneous remission," the group as a whole does not.

3. Our results fail to support the view that therapy starts a process of change which continues after the interviews cease. Therapy gains remain a year later, but they do not grow.

4. Despite the emotional dependence of children upon parents, therapy improvements occur without parent counseling.

5. Projective test improvements appear greater than those on an objective test.

6. Since our projective test is not a validated instrument, it may be a poorer index than the objective test, which has known validity.

7. Effective therapy can be done in a school setting, insofar as tests may measure outcomes.

8. Sex differences are notably absent.

9. Within the limits of our procedures, therapy gains are not proportionate to the number of sessions.

10. No general claims for client-centered therapy can be made from this study since all children were seen by the same therapist.

11. Owing to the absence of behavioral data, we do not know whether test improvements reflect actual changes in life adjustments.

SUMMARY OF SCORING METHOD AND RELIABILITY OF THE SENTENCE COMPLETION TEST

Theme Analysis

Each completion is assigned to one of 12 content categories, chosen to cover the main areas of living of concern to children—work, play, health, the future, the self, and persons important in the child's life. Category 7, Animals, is included because a random sample of 50 records drawn from the research group's schoolmates showed frequent mention of animals. The list of themes follows.

- | | |
|-------------|--------------------|
| 1. Self | 7. Animals |
| 2. Siblings | 8. School |
| 3. Mother | 9. Recreation |
| 4. Father | 10. Sickness-Death |
| 5. Family | 11. Future |
| 6. Others | 12. Miscellaneous |

The principles governing categorization, definitions of themes, and scoring samples are given in detail in the original manual (7, pp. 120-188).

Attitude Analysis

Each completion is classified as Positive, Negative, or Neutral, depending upon its affective tone. This is independent of the theme category. That is, a completion may, for example, be placed in a theme category which ordinarily implies a positive attitude and yet be given a Negative attitude score.

Adjustment Rating Analysis

This consists of making judgments on 10 characteristics, each of which is evaluated on a seven-point scale. Each scale represents a continuum from best to worst adjustment, with higher ratings meaning worse adjustment. The 10 qualities are chosen to include only those apt to show change if therapy occurs, and all to tap some aspect of personal adjustment. An eleventh index, Mean Adjustment Rating, is calculated for each record by averaging ratings on the 10. This is our index of general adjustment status. The list of 10 qualities, together with abbreviated definitions, is given below.

1. Anxiety—a feeling of painful uneasiness regarding the future, characterized by mingled or alternating dread and hope (30, p. 16).
2. Security—a condition of present safety, assurance, comfort, and being loved.
3. Dependence—a condition of being unable to sustain or help oneself, or to perform a given act, without the consent or aid of some external force, person, or group.

4. Conflict—an unpleasant emotional state resulting from opposed tendencies or desires, each of which is of approximately equal strength.

5. Affectivity—the disposition to emotional experience, the capacity to be moved to feelings by internal or external events.

6. Flexibility—the capacity to alter one's attitudes or behavior in adaptation to changing circumstances.

7. Spontaneity—the capacity to initiate behavior on the basis of inner feeling, proneness, or temperament, as contrasted with the need for constraint or immediate external influence.

8. Self-regard—the general disposition to react with approval or disapproval toward oneself.

9. Attitude Toward People—the general disposition to react with approval or disapproval toward persons, exclusive of family members.

10. Family Attitude—the general disposition to react with approval or disapproval toward the family unit, as well as its members.

In the original scoring manual (7, pp. 120-188), each of the above qualities is defined in greater detail, both in general and in terms of its extremes. A list of Sentence Completion Test characteristics which may contribute to the rating also accompanies each definition. Likewise, each of the seven scale positions is defined.

Reliability

All reliability determinations are based upon the same 10 records, randomly drawn from the control and experimental cases, except where otherwise noted.

Split-Half Reliability

The test is divided into halves by matching items for presumed Attitude stimulus value. For each half of each protocol, we total the number of completions falling into each of our three Attitude categories. Chi square tests the hypothesis that the halves are samples from a common population. The 10 individual chi squares (one for each of 10 records) are summed to give $\Sigma \chi^2 = 15.92$, $p < .80$, thus affording no basis for rejecting the hypothesis. Pooled $\chi^2 = 0.40$, $p < .90$, and Interaction $\chi^2 = 15.52$, $p < .70$, so that the pooled figure is an adequate estimate (27, pp. 188-192). The test halves are probably alike in Attitude categories elicited.

The split-half analysis is not made for Themes, since the test has no equivalent halves for them. It is also irrelevant to the Adjustment Ratings, which do not depend on the same test items in all records.

TABLE 9
INTERCORRELATIONS OF SENTENCES ADJUSTMENT RATING SCALES AT
FIRST TESTING OF 34 SUBJECTS^a

Scales	Anx- iety	Secur- ity	Depend- ence	Con- flict	Affec- tivity	Flexi- bility	Spon- taneity	Self	People	Family
Aspects of Adjustment										
Security	.80									
Dependence	.50	.45								
Conflict	.45	.50	.49							
Affectivity	.44	.36	.27	.26						
Flexibility	.15	.07	.01	.02	.02					
Spontaneity	.32	.23	.47	.30	.09	.19				
Self	.68	.60	.40	.39	.29	.25	.30			
People	.42	.29	.17	.47	.31	.07	.15	.45		
Family	.42	.43	.10	.34	.50	.22	.01	.43	.36	
Over-All Adjustment										
Mean Rating ^b	.75	.69	.51	.59	.48	-.09	.36	.62	.50	.43

^a For $df=32$, the following r values are required at the given p levels: (a) for $p<.001$, $r=.54$; (b) for $p<.01$, $r=.44$; (c) for $p<.02$, $r=.40$; (d) for $p<.05$, $r=.34$.

^b Corrected for contamination by McNemar's formula for r between a subtest and a total test minus that subtest (17, p. 139).

Intercorrelations of Rated Adjustment Characteristics

Although the scoring manual (7, pp. 120-188) directs separate consideration of each characteristic, it is possible that some global quality is being rated under different names. To test this, we intercorrelate ratings on all the scales, by combining the 17 experimental and 17 control cases at first testing, to give $N = 34$. Table 9 gives the Pearson r 's, and shows that 22 of the 45 interscale r 's (excluding Mean Rating) are significant at the .05 level or better ($r \geq .34$), so that we surely do not have 10 discrete measures. Our aim was not to predict changes on these individual scales, but to use them to derive a general notion of adjustment status, that embodied in the Mean Rating. Table 9 shows that 9 of the 10 scales correlate with Mean Rating at the .05 level or better, indicating internal consistency. Thus, the general adjustment index is based upon converging measures. While we might have used a direct assessment of over-all adjustment instead of Mean Rating, previous research with a similar test showed a global rating to be less reliable and valid than a score derived from specific items (31).

Rescoring Reliability

The writer rescored 10 records after six months. Agreement is 95% for Themes, whereas chance expectancy is 8%, since 12 categories are involved. For Attitudes, the agreement is 93%,

while chance expectancy is 33%, since three categories are involved. Both of the obtained intra-judge agreements are significant beyond the .001 level.

For the adjustment rating analysis, the Pearson r between the two sets of pooled ratings is .91, significant at the .001 level. On the 10 individual scales, rho ranges from .69 to .996, and all are significant at or beyond the .05 level. Thus, all the scales can be reliably used by the single scorer.

Interjudge Reliability

To check the experimenter's work, two outside judges evaluated the same 10 cases. Their training for the task consisted of study of the scoring manual, which includes a demonstration analysis of one case. Then each judge scored two sample cases as a training exercise. The sample analyses were carefully studied by the investigator who then gave detailed criticisms to each judge individually.

The work of three judges is compared. Judge A, the writer, had worked intensively with disturbed children, and had been trained in other projective tests. Judge B, a graduate student, was mainly an adult therapist, but had used non-projective tests with children. Judge C, a clinical psychologist, had diagnostic and therapeutic experience with children and adults, but projective test experience with adults only. Thus, all judges had clinical experience with children, but with methods other than sentence completions.

Based on 10 records, interjudge agreement on Themes ranges from 76% to 86%; agreement on Attitudes ranges from 80% to 86%. These percentages differ significantly from chance expectancy at the .001 level. Pearson r 's for the pooled Adjustment Ratings range from .64 to .78, all significant at the .001 level. Agreement on individual scales is tested by the W statistic, which ranges from .53 to .89, significant at the .02 level or better for 8 of the 10 scales, the 2 exceptions being Dependence and Flexibility. For Mean

Rating, our general adjustment index, $W = .82$ ($p < .01$).

The disagreement on Dependence and Flexibility is general, and is not a function of any particular pair of judges. Rho coefficients show absence of correlation between any pair of judges. Since Judge A , the investigator, could reliably use these scales, lack of interjudge agreement may reflect inadequacies in the scoring manual.

SUMMARY OF SCORING METHOD AND RELIABILITY OF THE FOLLOW-UP LETTERS

Therapy Memories

Each letter is classified Positive, Negative, or Neutral, depending upon its affective tone.

1. Positive—expressions of pleasant memories, fun, benefit derived, or that the therapist is missed.
2. Negative—expressions of displeasure about the sessions, statements denying their value, or hostile demands for explanations of their purpose.
3. Neutral—mere recitals of recalled activities, without statements implying approval or disapproval.

Reports of Present Status

Each letter is judged Positive, Negative, or Neutral for reported life status, depending upon the degree of well-being or personal satisfaction expressed.

1. Positive—statements that the child is doing well, whether in terms of personal happiness or school progress; favorable references to currently enjoyed activities.
2. Negative—accounts of failure experiences, unhappiness, or generally poor progress.
3. Neutral—noncommittal remarks about current activities, as well as complete omission of reference to present life status.

Reliability

All reliability determinations are based upon the total group of 15 cases followed up.

Rescoring Reliability

The writer analyzed each letter twice, with a six-month interval between. Intrajudge agreement is 100% for therapy memories and 93% for reports of present status, while chance expectancy is 33%, since three categories are involved. The obtained agreement differs significantly from chance expectancy at the .001 level.

Interjudge Reliability

Three judges independently classified all letters, working from the scoring manual (7, pp. 189-192). Judge A is the writer. Judge B is a clinical psychologist with four years of diagnostic and therapeutic experience with children and adults. Judge C is a graduate student in the humanities, without prior experience of this kind. Agreement between pairs of judges ranges between 87% and 93% on therapy memories, and reported status agreements are all 80%. All differ significantly from chance expectancy at the .001 level. Judges B and C have identical agreement percentages with the writer, suggesting that the task of judging this material does not require training beyond that given in the scoring manual.

REFERENCES

1. AXLINE, VIRGINIA M. *Play therapy*. Boston: Houghton Mifflin, 1947.
2. AXLINE, VIRGINIA M. Play therapy experiences as described by child participants. *J. consult. Psychol.*, 1950, 14, 53-63.
3. BARRON, F., & LEARY, T. F. Changes in psychoneurotic patients with and without psychotherapy. *J. consult. Psychol.*, 1955, 19, 239-245.
4. BILLS, R. E. Nondirective play therapy with retarded readers. *J. consult. Psychol.*, 1950, 14, 140-149.
5. CARTWRIGHT, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, 19, 357-363.
6. DORFMAN, ELAINE. Play therapy. In C. R. Rogers, *Client-centered therapy*. Boston: Houghton Mifflin, 1951. Pp. 235-277.
7. DORFMAN, ELAINE. Personality outcomes of client-centered child therapy. Unpublished doctoral dissertation, Univer. of Chicago, 1955.
8. DYMOND, ROSALIND F. Adjustment changes over therapy from self-sorts. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 76-84.
9. DYMOND, ROSALIND F. Adjustment changes over therapy from Thematic Apperception Test ratings. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 109-120.
10. EDWARDS, A. L. *Statistical analysis for students in psychology and education*. New York: Rinehart, 1946.
11. FLEMING, LOUISE, & SNYDER, W. U. Social and personal changes following non-directive group play therapy. *Amer. J. Orthopsychiat.*, 1947, 17, 101-116.
12. GOLDBERG, SHARON. Quantitative comparison of verbal and graphic expressions of eight personality variables. Unpublished master's paper, Univer. of Chicago, 1950.
13. GRUMMON, D. M. Personality changes as a function of time in persons motivated for therapy. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 238-255.
14. KENDALL, M. G. *Rank correlation methods*. London: Charles Griffin, 1948.
15. LANDISBERG, SELMA, & SNYDER, W. U. Non-directive play therapy. *J. clin. Psychol.*, 1946, 2, 203-214.
16. MACHOVER, KAREN. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.
17. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
18. PUGH, R. W. An investigation of some psychological processes accompanying electric convulsive therapy and non-directive psychotherapy with paranoid schizophrenia. Unpublished doctoral dissertation, Univer. of Chicago, 1949.
19. READER, NATALIE. An investigation into some personality changes occurring in individuals undergoing client-centered therapy. Unpublished doctoral dissertation, Univer. of Chicago, 1948.
20. ROGERS, C. R. Measuring personality adjustment in children nine to thirteen years of age. *Teach. Coll. Contrib. Educ.*, 1931, No. 458.
21. ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.). *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
22. ROHDE, AMANDA R. *Sentence completions test manual*. New York: Author, 1947.
23. ROTTER, J. B., & WILLERMAN, B. The incomplete sentences test as a method of studying personality. *J. consult. Psychol.*, 1947, 11, 43-48.
24. SEEMAN, J. A study of the process of non-directive therapy. *J. consult. Psychol.*, 1949, 13, 157-168.
25. SEEMAN, J., & EDWARDS, B. A therapeutic approach to reading difficulties. *J. consult. Psychol.*, 1954, 18, 451-453.
26. SHOR, J. Report on a verbal projective technique. *J. clin. Psychol.*, 1946, 2, 279-282.
27. SNEDECOR, G. W. *Statistical methods*. Ames, Iowa: Iowa State Coll. Press, 1946.
28. TAYLOR, J. W. Relationship of success and length in psychotherapy. *J. consult. Psychol.*, 1956, 20, 332.
29. TEUBER, H. L., & POWERS, E. Evaluating therapy in a delinquency-prevention program. *Proc. Ass. Res. nerv. ment. Dis.*, 1951, 31, 138-147.
30. WARREN, H. C. *Dictionary of psychology*. Boston: Houghton Mifflin, 1934.
31. WISCHNER, G. J., ROTTER, J. B., & GILLMAN, R. D. Projective techniques. In S. W. Bijou (Ed.), *The psychological program in AAF convalescent hospitals*. Washington: U. S. Government Printing Office, 1947. Pp. 150-157.

(Accepted for publication February 4, 1957)



Psychological Monographs: General and Applied

Consistency of Projective Movement Responses¹

JOSEPH F. OKARSKI

Teachers College, Columbia University²

THERE has been increasing emphasis in the literature to the effect that the contradictory results encountered in validation studies of projective techniques may be due partly to the possibility that some of the underlying assumptions are sound but others faulty. For this reason it has been urged that closer scrutiny of

specific practices and their theoretical bases is needed to screen those which may be retained from those which should be modified or discarded.

Among the practices which often have been challenged has been the observation and diagnostic interpretation of movement, as conceptualized in various symbolic or representational terms furnished by language, art, and other expressive media, and converted into psychological test procedures. The Rorschach "movement response," comprising those "concepts which projected some kind of action or life into ink blots" (13, p. 109), represents perhaps the best known application of the principle.

¹ This monograph is based on a doctoral dissertation at Columbia University. Laurence F. Shaffer, Chairman of the Doctoral Committee, and Edward V. Shoben, Jr. and Robert L. Thorndike, members of the committee, guided the development of the study as a whole. Helen M. Walker, Gordon F. Derner, and Herbert J. Zucker also contributed to the development of the study. Philip Brandis, Helen A. Diers, Joan Fox, Isidore Helfand, Shabse H. Kurland, Monroe L. Levin, Eugene Levitt, Reuben M. Margolis, Martha Hessel Page, Alfred Rubinstein, and Loron M. Simon helped in judging certain of the materials.

Special credit is due the United States Air Force which granted laboratory facilities and the cooperation of its staff under the Personnel Research Field Section Human Resources Research Center, Sampson Air Force Base, Geneva, New York, and which provided subjects for the experiment. Donald B. DuBois served as consultant on the scene.

The study is also indebted to the following: Weston A. Bousfield, who gave access to research facilities in the Department of Psychology at the University of Connecticut for important preliminary aspects of the study; The Psychological Corporation, which supplied the Levy Movement Cards; the late Oliver D. Havard, of Allentown, Pennsylvania, who gave financial support to certain phases of the study; and, finally, my wife, Ruth, who assumed many of the unenviable routine tasks and other burdens involved.

I am deeply grateful to all.

² Now at the Rochester Guidance Center, New York.

Fundamental to the observation and interpretation of conceptualized movement in the projective techniques has been the explicit or implicit assumption that the movement response occurs as a function of something deep rooted and stable in the individual. Observed as a test response, it has thus been presumed to yield revealing and dependable clues to personality and behavior. While some evidence directly or indirectly relevant to this proposition is to be found in the experimental literature, research to date has not been conclusive. The present study, therefore, has sought to add some clarification.

Methodological Approaches

Previous approaches to the problem have usually taken one of two directions:

No-4, 5, wanting

(a) attempting to establish a direct relationship between movement responses and personal attributes thought to have important and durable psychological significance and (b) examining the consistency or inconsistency of movement scores and derived indices. The search for correlates has been inconclusive partly because validity of the criteria used to define correlative variables has not always been demonstrated. More basically, however, these studies have been weak because they have uncritically assumed the movement indices provided personally relevant behavioral data largely insusceptible to chance distortion or to systematic error. Since both experimental successes and failures might conceivably be accounted for in either of these terms, it has been difficult to determine in specific instances what interpretation might be put on observed results.

Studies following the second approach have in turn been indecisive because essential standards for testing clinical hypotheses concerning stability and variation in behavior have not been met. The shortcoming becomes evident when researches are weighed against basic requirements for examining the individual's variability in relation to his presumed "true" score or typical performance.

Zubin (29) has discussed these requisites at length. Key points may be applied to the present problem as follows: (a) the individual must be regarded as an independent universe of behavior; (b) each observation (in this case, the movement score) must be treated as a random sample from this universe; (c) sampling must be sufficiently extensive to provide a definitive estimate of the individual's range of variability within his universe; (d) observations must be based on a mathematical model (e.g., ex-

pansion of the binomial) which permits any observed range of variability to be tested rigorously for its significance; (e) the performance of the individual must be treated integratively as well as differentially in relation to that of others.

Studies of Consistency in the Movement Response

Researches pertaining to intra-individual variability in the movement response have usually been Rorschach studies. Most of them have been investigations of reliability using split-half, or single-repetition test-retest designs. In some cases special operations have been introduced to test the stability of the subject's scores under varied sets, suggestions, stresses, drugs, and other experimental conditions relating either to the subject, to the examiner, or to the test situation as a whole. Miller (17) has reviewed the work in this area and has outlined various possible sources of variation in the subject's scores.

Other considerations aside: the studies of reliability have generally provided no more than two estimates or tests of the subject's performance level. Assuming that the two estimates are independent and that there is equal probability for any two alternative performance levels (e.g., high or low) in either test, then the statistical significance of complete agreement, or of complete disagreement, between the two estimates is $.5^2 + .5^2$ or only .5—exactly what is expected by chance. Under these conditions it cannot be determined whether any subject has been significantly consistent or significantly inconsistent in his performance level.

Exceptions to this criticism have been longitudinal Rorschach studies like that done by Ives, Grant, and Ranzoni (11) and the study of internal consistency in

the Rorschach categories done by Wittenborn (28). Other considerations, however, have limited the implications of these studies for the present problem.

A number of researchers, for example, have suggested that, if the movement response truly involves significant personal attributes, it should reveal a broadly general character not confined to the Rorschach situation. Wittenborn (28) interpreted this hypothesis as consonant with underlying Rorschach assumptions. He suggested that, if the assumptions held, subjects might be expected to conceptualize movement with consistent frequency on a variety of tests which offered perceptual alternatives. Sanford (22) suggested that subjects who projected action into inkblots might be expected to inject action into scenes and situations which they depicted verbally. Others have likened the verb-adjective quotient used in analyzing verbal materials to the Rorschach movement-color ratio.

Going beyond the point of speculation, several investigators have offered presumably alternative or supplementary inkblot test forms, e.g., the Behn (31) and Harrower (8) series, claimed to yield results consistent with those of the Rorschach. Others have adopted or modified the Rorschach model for the experimental and diagnostic observation of movement in other types of tests involving interpretation or production of drawings, paintings, mosaic designs, toy arrangements, etc. The hypothesis of intra-individual consistency in conceptualizing movement might be extended to these techniques as well.

Several studies have offered direct evidence bearing on the consistency of the subject in movement production as observed in comparisons between different tests—the Rorschach has served as the

common denominator in these researches. The broadest was that of Mayman (15) who made comparisons between the Rorschach inkblots, on the one hand, and the Behn (31) and Harrower (8) series on the other. Extending beyond the inkblot frame of reference, King (12), and Allen, Ray, and Poole (1) compared Levy Movement card performance with that on the Rorschach. Hays, Gellerman, and Sloan (9) compared movement indices on the Rorschach with verb-adjective quotients in TAT protocols.

Since completion of the present study, Epstein, Nelson, and Tanofsky (7) have reported an exceptional investigation using ten sets of newly devised inkblots totaling 100 in all; apart from this, studies have been subject to the same weakness stressed in connection with the Rorschach studies described earlier: inadequate samplings of the individual's universe of test performance. Their results, therefore, have been equally indecisive. The question substantially remains: Does the movement response occur with such intra-individual consistency as to warrant the assumption that it represents some intrinsic, stable, personal variable? Resolution of the question would not only help to clarify the status of the movement response with respect to its basic assumptions in the Rorschach and similar tests, but might also contribute to the development of sound theory and practice in the projective techniques generally. The present study, therefore, has undertaken to explore the question according to some of the principles outlined by Zubin (29) and described earlier.

Hypothesis of the Present Experiment

It is hypothesized that, if the movement response, broadly defined, is a func-

tion of something deep rooted and stable in the individual, then his level of movement production as compared to that of other individuals is consistent over a range of different test situations.

PROCEDURES

As a general frame of reference, a movement response was defined as the verbal stipulation of some action. Subject to this limitation and to the modifications described below, conceptualized movement was observed in six different tests, all administered and scored by the experimenter. The tests were of two types conveniently designated as *Perceptual* and *Verbal*; three of each type were used.³

Perceptual Techniques

The perceptual tests included the Rorschach (20), Behn-Rorschach (31), and Levy Movement (21) cards, all administered and scored according to a standard procedure based on the Klopfer and Kelley (13) Rorschach manual. Besides the usual *M* (human), *FM* (animal), and *m* (inanimate) movement scores, a combined or sum movement (*SM*) score was obtained. All scores were converted to percentages to correct for individual differences in total number of responses per test and to provide a common base for comparison.

Verbal Techniques

The verbal techniques included two sentence construction procedures espe-

cially devised for the experiment. Stimulus words for these tests were chosen to provide cues minimally suggestive of response. It was postulated that, as the number of cues in the stimulus material was reduced, the likelihood of response to personal cues was increased. This procedure ostensibly gave greater assurance that response was more a function of the individual than of the stimulus material itself. The third verbal technique was a modification of Severn's (23) Heteronym Test. Brief descriptions follow.

Sentence Construction 1. The subject was asked to make up and write down as many short simple sentences beginning with the word *the* as he could in 30 minutes, using a "different idea" for each sentence. The test was presented as an experiment in "imagination," and the subject was encouraged to "stretch" his imagination as much as he liked and in any way he wished.

Sentence Construction 2. The subject was asked to compose one sentence for each of 40 different stimulus words like *the*, *and*, *but*, *for*, etc., incorporating these in any way he wished. Otherwise, instructions were substantially the same as those for Sentence Construction 1.

Heteronym Test. In the Heteronym Test the subject was presented visually with 40 two-syllable words, each having both verb and nonverb meanings, depending on the syllable stressed. Each was hidden in a matrix of scrambled letter mounted on a 4" × 6" card. The cards were arranged in four ordered sets including alternate cards containing nonheteronyms to mask the nature of the test. The cards were administered at intervals according to a predetermined plan (see *Administration of the Battery*). The subject was instructed to find the hidden word on each card and to say it aloud on doing so.

³ Only brief descriptions of the test procedures and materials are given here. Further details may be obtained by consulting Appendixes A-K, pp. 68-97, of the original publication; Okarski, J. F., Consistency of projective movement responses. Doctoral dissertation, Columbia University, 1956 (or Publication No. 17,072. Ann Arbor: University Microfilms).

Scoring

Sentence constructions. A scoring system similar to that used in the perceptual tests was experimentally developed. The method was based on adaptations of Rorschach scoring rules. A list of over 300 sample sentence constructions served as scoring criteria, e.g.: *M*—"The player made a home run." *FM*—"A cat just came in." *m*—"The tree was struck by lightning."

All sentence constructions were scored accordingly; raw scores were converted to percentages as in the perceptual tests.

The reliability of the scoring system was examined by submitting the responses of 64 subjects on Sentence Construction 1 to two independent judges, one scoring the odd numbered and one the even numbered items. Their scores were then compared with those of the experimenter. Ninety-one percent of the odd numbered items were scored with perfect agreement; 95% of the even numbered items were scored with perfect agreement.

Correlation coefficients comparing the scores given by the judges with those given by the experimenter were also computed. Corrected by Spearman-Brown formula, the coefficients ranged from .91 to .98. On the basis of these figures, the reliability of the evolved system of scoring appeared to be adequate for experimental purposes.

Heteronym Test. For the Heteronym Test, movement was scored in terms of the subject's pronunciation of the heteronym: if he pronounced the word according to the verb definition, he was credited with a movement response. The proportionate number of such pronunciations was recorded as the subject's movement score.

There appeared to be no basis for differentiating various subcategories of

movement as in the other tests; hence only the one index of movement was obtained. It was considered simply as a general indicator of movement response tendency corresponding most nearly, perhaps, with *SM* but suitable for comparison with any other type of score.

Since the technique was virtually self-scoring and quite objective, no test of scoring reliability was considered necessary.

Subjects

Complete data were obtained for 82 male subjects, 18 to 25 years of age. They were drawn from an experimental pool of military recruits in their sixth day of induction processing at Sampson Air Force Base, Geneva, New York. Subjects were required to be at least high school graduates; except for this limitation, selection was by random sampling procedures employed by the base for its own research purposes.

Subjects came from 17 different states throughout the eastern United States. Six were Negroes; the rest, white. Thirteen had up to two and one-half years of college training; 10 were college graduates. Only one reported having courses in psychology; these were of a rudimentary nature. No data concerning intelligence or other diagnostic attributes were obtained. Some notion of the social backgrounds of the subjects, however, might be inferred from the following information about parent occupation: 40% were skilled or semiskilled tradesmen or laborers in industry, business, mining, or construction; 23% were unskilled laborers; 18% held supervisory or managerial positions; 9% operated their own minor businesses; 5% were professional people; and 5% were farmers. Outside the background information given, subjects remained anonymous.

Administration of the Battery

Subjects were scheduled in pairs to complete the tests in one day. All tests except the Sentence Constructions were administered in individual sessions; the latter were given in joint sessions. Joint

TABLE 1
ODD-EVEN RELIABILITY COEFFICIENTS
FOR ALL SIX TESTS
(Based on percentage movement scores
for 82 Ss)

Test	Movement Category			
	<i>M</i>	<i>FM</i>	<i>m</i>	<i>SM</i>
Perceptual				
Rorschach	.34	.56	.25	.69
Behn	.41	.63	.42	.69
Levy	.59	.36	.49	.53
Verbal				
Sent. Constr. 1	.88	.70	.82	.87
Sent. Constr. 2	.75	.43	-.02	.74
Heteronym	—	—	—	.73

Note. — Coefficients corrected by Spearman-Brown formula.

and individual administrations were alternated to permit the subject suitable rest periods.

After a standardized orientation given to facilitate rapport and to dispel any examination set acquired by the subjects during the preceding days of classification and aptitude testing, the tests were given in the following order: (a) Heteronyms, first set; Rorschach; Heteronyms, second set; (b) Sentence Construction 1; (c) Heteronyms, third set; Behn-Rorschach; Heteronyms, fourth set; (d) Sentence Construction 2; (e) Levy Movement Cards. Before taking the tests subjects were given the option of withdrawing from the experiment without penalty in order to put participation on a voluntary basis. Brief preliminary interviews served to obtain personal data, as reported above, just before administration of the first test procedure.

Reliability of the Tests

Percentage scores on odd-even split-halves of each test were correlated with each other to obtain estimates of test reliability. In the case of the perceptual tests, scores were based on odd and even card groupings. In the verbal tests, scores were based on odd and even items.

Table 1 presents the values obtained, as corrected by Spearman-Brown formula. Coefficients varied widely, ranging from $-.02$ to $.88$. In general, the verbal tests appeared to be more reliable than

the perceptual. Sentence Construction 1 was the most reliable single test with values ranging from $.70$ to $.88$. The Rorschach was least reliable, with r running from $.25$ to $.69$.

Of the several movement indices, *SM* appeared to be the most reliable, r ranging from $.53$ to $.87$. Least reliable of the movement indices was *m*, with r running from $-.02$ to $.49$, excepting the single value, $.82$, obtained for Sentence Construction 1.

On the basis of these estimates, the techniques chosen for the experiment did not, as a whole, appear to be highly reliable. As reliability coefficients, however, the estimates presumed the measurement of a psychological variable which was relatively stable in character. It was precisely this feature, the stability of performance, which the experiment brought into question. It proposed that the answer to the question might be determined by observing the generality of the subject's performance among successive samples of behavior, as defined by each of several tests. The split-halves of each test, however, might also be regarded as successive samples or subsamples of behavior among which the generality of the subject's performance might be observed. The generality of the subject's performance within the limits given by split-half comparisons might thus be equally regarded as evidence concerning the relative stability or instability of the variable *within* each test.

Treatment of the Data

It was assumed that a particular form of behavior represents something deep rooted and stable in the individual if it occurs repeatedly in different situations with such regularity as to minimize chance occurrence. Treatment of the data, therefore, hinged on some method

of classification which permitted probability calculation.

Classification of scores and probability. All percentage scores were first classified as high or low, depending on whether they fell above or below their respective medians as shown in Table 2. Scores falling close to the median were computed to three or more decimal places to eliminate ties and determine classification. Where this procedure failed to break the distribution into equal halves, near-median scores were assigned at random to equalize the two class frequencies. Tied scores falling at the median in the larger of the approximate halves were serialized, and the number needed to fill out the smaller halves were chosen through a table of random numbers. The procedure was followed for all distributions otherwise failing of equal halves, except where the median fell at zero—in these cases, all zero scores were classified as low, and the rest as high.

Of the 21 distributions shown in Table 2, 12 required no random allocations. In the remaining distributions, the number of scores allocated at random from the larger half ranged from 1 to 4, with the average at 1.9.

Except for the *FM* category in the Levy test and Sentence Construction 2, where more than half of the subjects received zero scores, the classification procedure divided the sample in each movement category into two equal groups of high and low scoring subjects and defined equal probability (.5) for high and low scores in each test. In the *FM* category for the two tests just mentioned, proportions of high and low scoring subjects were $.37 + .63$ and $.22 + .78$ respectively, defining corresponding probabilities. Assuming each test to be an independent sampling of behavior, the prob-

ability of any high-low score pattern or combination in any category of movement could thus be computed either by expanding the binomial or by enumerating all combinations and permutations of high and low proportions. The procedure made simultaneously available (a) tests of significance for any amount of variation or consistency between high and low scores as shown by any one subject and (b) chance expectancy values against which the performance of the group as a whole could be evaluated.

Treatment of the data on the basis described involved certain dangers which should be noted. On the one hand, dichotomizing scores at the median sometimes defined wide limits within which the subject's scores might fluctuate and still be regarded as consistent as long as they were either above or below the median. On the other hand, if the subject's scores fluctuated about the median, they were automatically defined as inconsistent even though they might fluctuate within a narrower range than scores which otherwise qualified as consistent. The basis for treatment of the data thus raised the distinct possibility that consistent performance might not be detected in some cases and that individual differences in variation, as defined, might frequently represent artificial distinctions rather than true differences.

The method thus incurred serious risks of error. The importance of the experimental problem, however, appeared to justify some exploratory risks. In addition, in view of the way in which the data were distributed, it was questionable whether any other method was more appropriate. It was decided, therefore, to accept the risks involved while emphasizing that due caution might be necessary in interpreting the results.

TABLE 2

DISTRIBUTION OF ALL PERCENT MOVEMENT SCORES

(R: Rorschach; BR: Behn-Rorschach; LM: Levy Movement; S: Sentence Construction; H: Heteronyms; N=82)

Score	M					FM				
	R	BR	LM	S ₁	S ₂	R	BR	LM	S ₁	S ₂
.99-100										
.96-.98										
.93-.95										
.90-.92										
.87-.89										
.84-.86										
.81-.83										
.78-.80										
.75-.77			2							
.72-.74			0							
.69-.71			1							
.66-.68			1							
.63-.65			1							
.60-.62			0		2					
.57-.59			1		3					
.54-.56			1		3		2			
.51-.53			0		3		0			
.48-.50			4		4		1			
.45-.47			2	1	6		1			
.42-.44			2	1	6	1	1			
.39-.41			3	0	6	2	2			
.36-.38			3	0	6	3	5			
.33-.35	2		4	1	12	1	3			
.30-.32	0	1	2	0	1	4	5			
.27-.29	1	1	5	1	4	5	6			
.24-.26	2	0	6	1	10	12	7			
.21-.23	4	0	4	3	6	8	5			
.18-.20	8	7	12	5	7	11	12	2	1	
.15-.17	9	8	1	8	2	7	9	0	0	
.12-.14	10	11	9	5	1	4	8	5	1	
.09-.11	7	7	7	13		8	3	4	2	
.06-.08	16	19	3	21		3	4	9	4	1
.03-.05	8	12	1	14		6	3	10	38	17
.00-.02	15	16	7	8		7	5	52	36	64
$\bar{X}$	.11	.09	.25	.10	.35	.18	.22	.03	.03	.01
Mdn	.09	.07	.21	.08	.35	.19	.19	.00	.02	.00
S	.08	.07	.18	.09	.12	.11	.12	.05	.03	.02

Plan of comparison. If the experimental hypothesis were generally true, it would be expected that for any category of movement all subjects, or a significantly large majority, would score either high or low on all six tests with a regularity beyond that anticipated on the basis of chance.

If the hypothesis were not generally true, then it would be expected that subjects would shift between the two performance levels, with high-low score combinations showing either uniformly

wide variation or individual differences distributed substantially according to chance.

To determine the status of the hypothesis, therefore, the high-low score patterns of all subjects were classified and proportionately distributed in each category of movement according to combinatorial order and corresponding probability. Chi-square tests (27) were then applied in each category to determine the nature of each distribution in relation to chance. The .05 confidence

TABLE 2 (continued)

Score	<i>m</i>					<i>SM</i>					
	R	BR	LM	S ₁	S ₂	R	BR	LM	S ₁	S ₂	H
.99-.100								1			
.96-.98								0			
.93-.95								0			
.90-.92								0			
.87-.89								1			
.84-.86								0			1
.81-.83								0			0
.78-.80								0			2
.75-.77						1		1			2
.72-.74						0		0			3
.69-.71						0		2			6
.66-.68						1		2			6
.63-.65						1	2	0		1	12
.60-.62						1	2	2		3	2
.57-.59						2	1	1		5	5
.54-.56						1	3	1	1	3	6
.51-.53						2	4	1	1	4	11
.48-.50						3	5	6	2	6	8
.45-.47						4	4	4	0	6	4
.42-.44						5	7	4	4	6	6
.39-.41						5	5	4	5	3	3
.36-.38				1		7	2	6	3	9	2
.33-.35				0		8	8	8	4	7	2
.30-.32				1		8	11	4	5	8	1
.27-.29				0		3	6	8	5	4	
.24-.26			3	2		4	6	8	10	8	
.21-.23			1	5		7	1	4	12	3	
.18-.20		1	5	3		6	1	6	6	4	
.15-.17		0	4	10		4	2	1	6	1	
.12-.14	2	1	9	19		1	4	2	7	1	
.09-.11	5	5	9	10		1	1	1	4		
.06-.08	12	17	8	19	4	2	3	0	3		
.03-.05	21	23	8	6	39	2	1	0	2		
.00-.02	42	35	35	6	39	3	3	4	1		
$\bar{X}$	.03	.04	.07	.12	.02	.32	.34	.36	.25	.38	.57
<i>Mdn</i>	.02	.03	.06	.10	.02	.32	.33	.33	.23	.38	.55
<i>S</i>	.04	.04	.08	.07	.02	.16	.16	.18	.12	.12	.12

level was used to define the boundary between a chance and a nonchance distribution.

Supplementary analyses were undertaken to detect any significant trends and relationships in performance over the three verbal tests, the three perceptual tests, and all possible pairs of tests.

Finally, the relation of the subject's variability in one category of movement to that in others was examined.

Subject sampling. Comparisons were made for two alternate groups of sub-

jects. The first comprised all 82 subjects as described earlier. The second comprised a subsample of 64 of these same subjects who gave a minimum of 15 Rorschach or 15 Behn-Rorschach responses. The purpose of the alternative sampling was to observe the effect of error due to the use of percentage movement scores based on small response totals.

RESULTS

It is hypothesized that, if the conception of movement is a function of some-

TABLE 3

SIX-TEST PERFORMANCE: PROPORTIONATE DISTRIBUTION OF HIGH-LOW MOVEMENT SCORES OBSERVED AND EXPECTED BY CHANCE IN EACH MOVEMENT CATEGORY FOR ALL SUBJECTS
($N=82$)

Move- ment	Proportionate frequencies in each category of movement					
	Probable		Observed			
	$M,$ $m,$ SM	FM	M	FM	m	SM
6-0	.016	.005	.024	.000	.037	.037
5-1	.094	.046	.122	.061	.134	.122
4-2	.234	.167	.256	.207	.195	.207
3-3	.313	.303	.232	.280	.280	.256
2-4	.234	.297	.195	.244	.183	.232
1-5	.094	.150	.134	.171	.122	.110
0-6	.016	.031	.037	.037	.049	.037
χ^2 Sig. ^a			7.25 .50	2.84 .90	11.94 .10	6.63 .50

^a Six degrees of freedom.

thing deep rooted and stable in the individual, then his level of movement production, as compared with that of other individuals, is consistent over a range of different test situations. Analysis of the data gave no support to the hypothesis as a general proposition, as shown by the results below.

Six-Test Performance

Table 3 shows the results for all 82 subjects on the full test battery. Given is the proportionate distribution of high-low score patterns probable and observed in each movement category, with the derived chi square and significance values. Individual patterns varied from constantly high (6-0) to constantly low (0-6). Only at these extremes could the subject's performance be regarded as significantly consistent at the .05 level or better. Less than 4% of the subjects in any movement category scored constantly high; less than 5% in any instance scored constantly low. Distribution chi squares

TABLE 4

SIX-TEST PERFORMANCE: PROPORTIONATE DISTRIBUTION OF HIGH-LOW MOVEMENT SCORES OBSERVED AND EXPECTED BY CHANCE IN EACH MOVEMENT CATEGORY FOR THE SUBSAMPLE
($N=64$)

Move- ment	Proportionate frequencies in each category of movement					
	Probable		Observed			
	$M,$ $m,$ SM	FM	M	FM	m	SM
6-0	.016	.007	.031	.000	.047	.047
5-1	.094	.057	.094	.078	.156	.125
4-2	.234	.185	.297	.250	.203	.203
3-3	.313	.309	.250	.266	.297	.266
2-4	.234	.282	.172	.203	.156	.219
1-5	.094	.135	.125	.172	.109	.109
0-6	.016	.026	.031	.031	.031	.031
χ^2 Sig. ^a			5.60 .50	4.98 .75	9.82 .25	6.62 .50

^a Six degrees of freedom.

were not significant at the .05 level, indicating no great discrepancy between observed and chance frequencies.

Table 4 shows a corresponding analysis for the subsample of 64 subjects whose Rorschach and Behn-Rorschach records showed response totals of 15 or more. High-low score patterns were distributed in substantially the same way as those for the larger sample. Distribution chi squares were again insignificant at the .05 level.

Both samplings thus showed chance distribution of individual differences in variation between high and low movement scores. In these terms, it was apparent that, if the hypothesis held, it did so for a few cases only and not for the great majority of subjects. Furthermore, since frequencies were distributed generally according to chance, there was no assurance that even these few cases were not accidental.

It was also apparent from the two subject samplings that the results of the an-

alysis were not seriously affected by either the inclusion or exclusion of subjects with low response totals. Succeeding analyses revealed similarly duplicating results for the two samples; for this reason, and because the samples were not otherwise independent, only the results for the larger group are hereafter presented.

To determine whether the few cases who seemed to be consistent represented the hypothesized phenomena, the data for the six tests were analyzed further to detect any supportive trends. For this purpose the high-low score combinations of all subjects were redistributed according to the degree of variation or maximum intertest agreement and disagreement represented. In effect, this procedure merely served to combine distribution frequencies for corresponding extreme intervals as shown in Table 3. Thus, proportionate frequencies for intervals 6-0 and 0-6 were combined into a single interval representing six tests in agreement; frequencies for intervals 5-1 and 1-5 were combined in a single interval representing five tests in agreement, etc. Chi-square tests were then applied as before.

Only in the *m* category was a trend in the hypothesized direction observed. Roughly 9% of the subjects, as compared to an expected 3%, were significantly consistent with all six tests in agreement on movement score level. About 26%, as compared to an expected 19%, showed five tests in agreement. The remaining proportions fell below the expected values. The distribution chi square was 11.46, significant at the .001 level. Thus, there was no support for the hypothesis except in *m* production.

For a still coarser analysis the number of subjects in each movement category who were significantly consistent (six

tests in agreement) were compared with the number who were not significantly consistent (three to five tests in agreement). This involved combining proportionate frequencies for intervals 6-0 and 0-6, as shown in Table 3, in one cell and combining frequencies for the remaining intervals in a second cell. Chi square was then computed to determine whether the observed number of significantly consistent subjects exceeded chance expectancy. Again only in *m* were the results significant. Chi square with Yates' correction was 6.2 with a probability of .02.

On the basis of the coarser analysis, the hypothesis appeared to be confirmed for a small number of subjects on *m* production, and was rejected for the rest. It might be concluded that the conception of movement can be safely assumed to represent some intrinsic, stable variable for a small number of subjects only, and then only with respect to inanimate movement responses.

Supplementary Analyses

It was apparent from the previous results that the subject's movement score generally varied from test to test. A question thus arose as to the roles played by specific tests or types of tests in determining the outcome of the experiment. Analysis was therefore narrowed to the areas of performance defined by the three verbal tests, the three perceptual tests, and all possible pairs of tests.

Three-Test Performance

The high-low score patterns for each set of three tests were examined exactly as were the data for the six tests. Only for the perceptual tests were the results significant, as shown in Table 5 which gives the proportionate high-low score distribution in each movement category.

About 17% to 22% of the subjects

scored constantly high; about 20% to 23% scored constantly low. All these proportions were greater than chance values, while those for intervening intervals were lower. Distribution chi squares were all significant at the .01 level or better. It was thus apparent that, for every movement category, significantly large numbers of subjects were consistent in level of movement production on the perceptual tests.

Combining distribution values according to the maximum intertest agreement represented by high-low score combinations, it may be seen that the probability for consistent performance in any one movement category was .25. In other words, perceptual test samplings did not provide a crucial test as to the significance of any one subject's performance. Hence, while significantly large numbers of subjects were consistent in these tests, it could not be said that any subject was *significantly* consistent at the standard .05 confidence level.

Trends in the hypothesized direction were nevertheless evident for every movement category. In this respect, the results for the perceptual tests contrasted with those for the verbal tests. It might be concluded that performance in these tests gave more support to the hypothesis of consistency in movement production and that the conception of movement was more likely to be a function of something deep rooted and stable in tests of this character, perhaps, than in those of the verbal type. It appeared unlikely, however, even on these terms, that the hypothesis could hold equally for all or even a majority of subjects.

Paired-Test Comparisons

Using four-fold contingency tables, phi coefficients and chi squares were

TABLE 5
PERCEPTUAL TESTS: PROPORTIONATE DISTRIBUTION OF HIGH-LOW MOVEMENT SCORES OBSERVED AND EXPECTED BY CHANCE IN EACH MOVEMENT CATEGORY
($N=82$)

Movement	Proportionate frequencies in each category of movement					
	Probable		Observed			
H-L	<i>M, m, SM</i>	<i>FM</i>	<i>M</i>	<i>FM</i>	<i>m</i>	<i>SM</i>
3-0	.125	.091	.220	.183	.171	.220
2-1	.375	.341	.256	.232	.390	.280
1-2	.375	.409	.329	.354	.207	.280
0-3	.125	.159	.195	.232	.232	.220
χ^2 Sig. ^a			12.63 .01	13.83 .005	15.04 .005	15.63 .005

^a Three degrees of freedom.

computed to determine the interrelationship of high and low scores in each category of movement for all possible pairs of tests. Table 6 shows the results. Coefficients ran uniformly low, ranging from .41 to $-.27$ with 19 out of 60, or nearly a third, on the negative side. All together, only 9 showed significant interrelationships in performance.

Significant phi values were largely confined to the perceptual tests, as might have been expected from the results for three-test comparisons. Performance on the Rorschach was significantly related to that on the Behn-Rorschach in every movement category and to that on the Levy Movement Cards in the categories *M* and *m*. In contrast, there were no significant relationships between the Levy and the Behn, though all phi values were positive. There were no significant relationships between any two verbal tests and few for verbal-perceptual pairings.

It is notable that, while the only significant result observed in six-test performance was in *m*, paired-test compari-

sons showed most of the nine significant phi values in *M*. Included were the only significant relationships observed in verbal-perceptual comparisons; one of these, however, was significantly negative: $-.27$ for the Levy vs. Sentence Construction 2.

It may be concluded from these and from the foregoing results that, if the conception of movement bears any significant relationship to psychological processes in the individual for more than a minority of subjects, then the relationship is largely specific to certain tests or types of tests and that performance is not generally predictive from one type of test to another for any individual.

Intercategory Comparisons

Results of the experiment showed the presence of individual differences in variation between high and low movement scores. These differences were so distributed that the hypothesis was, in general, only tenuously upheld for some subjects and not for others *within* each movement category. Two further questions thus suggested themselves: (a) If movement scores were not themselves generally stable so as to justify the assumption that something deep rooted and stable was reflected, was there anything to suggest that something of an intrinsically stable character was represented by individual differences in variation? (b) Since the experimental hypothesis was upheld (even though tenuously) for some subjects in each movement category, was there anything to suggest selective consistency with possibly stable individual differences generally defined on the basis of the specific category or categories in terms of which consistent tendencies were selectively expressed?

Comparisons *between* the several

movement categories were undertaken in the hope that some answers to these questions might be obtained.

Individual Differences in Variation

It is hypothesized that, if individual differences in variation between high and low movement scores represent intrinsically stable characteristics, then the degree of variation shown by the subject in one category of movement is consistent with that shown in other categories of movement.

To examine the hypothesis, the data were first treated as in paired-test comparisons, except that movement categories instead of tests were paired, and corresponding individual differences in variation, instead of classified movement

TABLE 6

PAIRED TESTS: PHI VALUES FOR PAIRED-TEST COMPARISONS

(R: Rorschach; BR: Behn-Rorschach; LM: Levy Movement Cards; S₁: Sentence Construction 1; S₂: Sentence Construction 2; H: Heteronyms; N=82)

Paired tests	Movement category			
	<i>M</i>	<i>FM</i>	<i>m</i>	<i>SM</i>
Verbal				
S ₁ -S ₂	-.02	.00	.12	.07
S ₁ -H	-.02	-.12	.02	-.07
S ₂ -H	-.12	.06	-.07	-.17
Perceptual				
R-BR	.31*	.31*	.22*	.41*
R-LM	.22*	.15	.22*	.17
BR-LM	.12	.20	.17	.17
Verbal-Perceptual				
S ₁ -R	.02	.02	.12	.07
S ₁ -BR	.31*	.07	.12	.17
S ₁ -LM	.07	-.15	-.02	-.02
S ₂ -R	-.07	-.06	.17	-.12
S ₂ -BR	.02	-.12	.07	-.17
S ₂ -LM	-.27*	-.10	.07	.17
H-R	.17	.12	.02	.17
H-BR	.27*	-.12	.02	.07
H-LM	.02	-.06	.07	.12

* Significant at the .05 level as indicated by chi square with one degree of freedom.

TABLE 7

INTERCATEGORY COMPARISONS: PROPORTIONATE DISTRIBUTION OF SUBJECTS, PROBABLE AND OBSERVED, ACCORDING TO THE NUMBER OF MOVEMENT CATEGORIES—*M*, *FM*, OR *m*—IN WHICH THE SUBJECT WAS RELATIVELY CONSISTENT ON THE FULL BATTERY OF TESTS ($N=82$)

No. of categories	Proportionate frequencies	
	Probable	Observed
3	.029	.073
2	.198	.122
1	.444	.451
0	.328	.353
χ^2		8.03
Sig.		.05 ^a

^a Three degrees of freedom.

^b One degree of freedom.

scores, were plotted against each other in contingency tables. Individual differences in variation were defined by the maximum interest agreement shown in high-low score combinations. Chi squares were then computed to determine whether degree of variation in one category of movement was significantly related to that in others, for the verbal tests, the perceptual tests, and the full battery.

Chi squares were significant for *M* to *SM* comparisons only. Excepting these two categories, therefore, there was no indication that individual differences in variation were significantly related from category to category. The hypothesis was thus rejected as a general rule. Inspection of raw score data indicated that the connection between *M* and *SM* indices was largely a function of part-whole score relationship and so offered no exception to this conclusion.

The nature and effect of the part-whole score relationship was most clearly evident in the Heteronyms, Sentence Construction 2, and Levy Movement

tests. In the Heteronym Test, *FM* and *m* responses were not differentiated; only the general, *SM* score was obtained. Since the *SM* score was entered into all comparisons, the subject's *M* and *SM* scores were necessarily identical. Similar situations obtained for the Sentence Construction 2 and Levy tests because relatively few *FM* and *m* responses were evoked. Since the subject's *M* and *SM* scores were thus identical—or nearly so—in three of the six tests, individual differences in variation for both movement categories were heavily biased toward duplication. The significance of the relationship between them, therefore, was more indicative of limitations in the test procedure than of intrinsically stable characteristics in the individual.

Extending the analysis, all subjects were classified and distributed according to the number of categories in which movement production level was relatively consistent. Since *SM* was not independent of all three common categories—*M*, *FM*, and *m*—it was excluded, which allowed a range of distribution from 0 to 3. For the full battery, performance was considered relatively consistent if five to six tests were in agreement; for both verbal and perceptual techniques, agreement among all three tests was required to meet the criterion. Exact probabilities were computed for each distribution interval by combining and permuting proportions observed to be relatively consistent, as defined above, within each category. A chi-square test of each distribution was then made to detect any significant discrepancy between probable and observed frequencies.

Only the full battery revealed any significant results. These are shown in Table 7, giving probable and observed distributions for both grouped and un-

grouped intervals. Observing the ungrouped data, it may be seen that about 7% of the subjects were, in fact, relatively consistent in all three movement categories, in comparison to a probable proportion of .029. The data thus showed that a limited number of subjects were, in fact, generally consistent in level of movement production beyond the likelihood of chance-determined performance at better than the .05 level of confidence.

The distribution chi square was 8.03, significant at the .05 level. Since the remaining observed proportions were relatively close to or below the expected values, it might be inferred that the number of subjects who were generally consistent was significantly greater than zero, even though small in absolute terms.

No other significant inferences could be drawn from these results. The hypothesis was thus tenable on a limited basis only. It might be concluded that individual differences in variation represented something of a stable character only for a small but significant number of subjects who were generally consistent in level of movement production. It might also be asserted that on the basis of these intercategory comparisons the limited and tenuous support found earlier for the original experimental hypothesis was confirmed and strengthened, and that the conception of movement in all three common categories, *M*, *FM*, and *m*, might well represent something deep rooted and stable for certain extreme cases, though not for the great majority of subjects.

Selective Consistency

It is hypothesized that if deep rooted and stable individual differences are ex-

pressed in terms of selective consistency, then all or a significantly large majority of subjects are relatively consistent in some one or more of the several movement categories.

It may be seen from the grouped data in Table 7 that a majority, or about 65% of the subjects, were relatively consistent in 1 to 3 movement categories; 35% were consistent in none. Chi square was .23, indicating that these proportions were almost exactly what was expected by chance.

There was thus no support for the hypothesis, even though a majority of subjects were consistent in one or more movement categories. Results obtained separately for the verbal and perceptual tests, not shown here, failed similarly to support the hypothesis. It was concluded that neither for the verbal, the perceptual, nor for the full battery of tests was it possible that intrinsically stable individual differences were defined in terms of selectively consistent tendencies favoring some one or more movement categories.

Summary of Results

The over-all performance of the subjects on the full test battery failed to support the hypothesis of consistency as a general proposition. Analysis showed the presence of widely distributed individual differences in variation between high and low movement scores for every category of movement. In these terms, the hypothesis was tenuously upheld, in most movement categories, for some few subjects only.

Some exception was apparent in *m* production where a small but significant number of subjects were significantly consistent. It was concluded that the conception of movement might safely be

assumed to represent something deep rooted and stable for a small number of individuals only and then only with respect to *m* production.

Intercategory comparisons of individual differences in variation, however, revealed a small but significant number of subjects who were generally consistent in all three common categories, *M*, *FM*, and *m*. On this basis, the above conclusion was somewhat strengthened and broadened to suggest that the conception of movement in all three categories might well represent something deep rooted and stable for certain extreme cases, though not for the great majority of subjects. Otherwise, individual differences in variation gave no evidence of intercategory stability so as to suggest that these differences might define something of a psychologically significant, stable character in the individual. There was also no evidence that stable individual differences might be defined in terms of selectively consistent tendencies favoring some one or more movement categories.

When trends and relationships among specific tests or types of tests were considered, positive results were confined almost exclusively to the perceptual tests. These showed trends in the hypothesized direction for every movement category. It was concluded that, if the conception of movement was a function of something deep rooted and stable in the individual, then it was more likely to be observed in tests of the perceptual than of the verbal type; but even on this basis, it could not be assumed to hold equally for all or even a majority of subjects.

Paired-test comparisons showed that trends among the perceptual tests were largely confined to significant relation-

ships between the Rorschach and Behn-Rorschach. All paired-test relationships, however—perceptual, verbal, and verbal-perceptual—were of a low order. It appeared that, if the conception of movement was related to something deep rooted and stable for more than a minority of subjects, then the relationship was largely specific to certain tests or types of tests and that performance could not be regarded generally predictive from one type of test to another for any individual.

DISCUSSION

Is the movement response generally stable? The answers suggested by previous studies might be summarized as follows: that the movement response is (a) generally stable, (b) generally unstable, (c) stable under certain conditions and unstable under certain conditions, (d) stable for some individuals or groups but not for others.

Under the more or less varied conditions represented by the different tests used in the present study, the experimental results might be said to agree that the movement response is *not* generally stable. They appeared to suggest, further, that the status of the movement response may vary depending on factors inherent in both subject sampling and the conditions under which the subject's movement responses are observed. The results might best be discussed, therefore, from the standpoints of the basic components of each of these sources of variation: (a) the individual and (b) the tests or types of tests used to sample his universe of movement responses. Some further implications may then be drawn in relation to general theory and practice in the projective techniques.

The Movement Response in Relation to the Individual

The results of the experiment suggested that when the individual is regarded as an independent universe of behavior and suitable tests of significance are applied to his performance over a range of different test situations, the conception of movement may well represent a significantly consistent variable for certain subjects. In these terms, the experimental results gave partial support to the predictions expressed by Wittenborn (28), Sanford (22), and others and encouraged the belief that the movement response as observed under a variety of different conditions or tests may well represent an intrinsically stable personal variable for some individuals or groups.

All three movement response categories—*M*, *FM*, and *m*—appeared to have personal relevance on these limited terms. The weight of evidence, however, was such as to emphasize most strongly that none of the commonly observed movement categories may be safely assumed stable and, hence, personally relevant for all individuals or groups in any or all test situations, including tests such as the Rorschach.

Question might be raised, however, as to whether the implication noted was not purely a function of the experimental method used. It may be recalled that a weakness of the method, as discussed earlier, was the possibility that consistently performing subjects might fail to be detected if their scores fluctuated about the median. It is entirely possible that in confirming the hypothesis of consistency for some subjects and not for others the experimental method merely pointed out those whose scores were, on the average, either more or less extreme

rather than more or less consistent. Due caution, thus, urged verification of the finding by further research.

While the experimental results did not foreclose the possibility that the movement response might be generally stable, they strongly discouraged such a possibility. In so doing, they followed a pattern common to a number of studies in which some effort—formal or informal—was made to examine either the stability or the validity of the movement responses and/or derived indices on the basis of the individual as well as group statistics. Milton (18), for example, found that both the *M* scores and the *M:C* ratios on the Rorschach were stable for only about a third of his subjects whether retested under experimentally induced set or under standard conditions. Using backward and forward test-retest Rorschach administration, Rabin and Sanderson (19) found the *M:C* ratios stable for 80% of their subjects and unstable for the rest. Thornton and Guilford (25) found split-half agreement for *M:C* for only 55% of their subjects. Altus and Thompson (3) used a standard administration of the Rorschach and found that, while consistent relationship between *M* scores and a test of intelligence held "almost invariably" for subjects whose movement scores were high, it failed to hold generally for those whose scores were low. The results of the present study, together with some of the more careful studies reported in the literature, thus strongly suggested that the movement responses and derived indices may be personally relevant for some individuals or groups but not for others and, further, that this conclusion may hold regardless of whether the movement responses are observed on an intra- or intertest basis and whether constant

or varied conditions are maintained.

It might also be asserted on the basis of the studies cited that, in confirming the hypothesis of consistency for some subjects but not for others, the general pattern of results yielded by the present study was not unique to its experimental approach. Thus, while due caution on the grounds described earlier strongly urged further verification, there appeared to be no reason why the findings should not be tentatively accepted.

Experimental and practical implications. If the findings are indeed correct, they may help to understand the conflicting and otherwise inconclusive implications generally reported on the basis of group results alone. For if the movement responses are stable or relevant for some individuals or groups but not for others, then, apart from any other sources of error, the positive or negative results observed in a given study may simply reflect the vagaries of unwittingly selective subject sampling. Considered in this light, the widely varied estimates sometimes reported for both the reliability and validity of the movement responses may not reflect as much on the doubtful nature of the evidence as on its sampling limitations.

The crucial question with respect to the reported findings may thus be not so much *whether* the movement responses provide dependable indices to personality and behavior but, rather, *for whom?* Practical considerations would thus seem to demand re-investigation of the findings under individual-oriented procedures to identify, if possible, those types of individuals or groups for whom the movement responses may safely be assumed relevant and to provide a discretionary basis for clinical interpretation.

The present study offered no basis for

testing any interpretive hypothesis such as might prove valid within the limits suggested. Those hypotheses in terms of which some significantly positive results have been reported, however, might be considered plausible: Altus and Thompson (3), Tucker (26), and others have found significant but low relationship between Rorschach movement indices and intelligence. Singer, Meltzoff, and Goldman (24), Meltzoff and Levine (16), and others found some significant relationship between Rorschach movement responses and motor inhibition, inner control, and fantasy. Altus (2) and others have reported positive relationship between Rorschach movement indices and evidences of maladjustment and psychopathology.

It remains to be established, however, for what classes of subjects these studies may offer dependable clinical hypotheses. From the nature of its subject sampling and the results of the present study, it might only be inferred that the movement responses given by subjects from some ostensibly normal or nonpsychiatric groups may rarely be found sufficiently consistent to warrant any kind of interpretation. It might be inferred further that, pending identification of the sampling characteristics or other convenient discriminative criteria which might define the psychologically significant populations, the clinician may have no assurance that any interpretation offered in a particular case actually relates to anything typical of the subject.

Indices of variation as diagnostic variables. A number of investigators have suggested that not only the consistency but the inconsistency and perhaps individual differences in amount of variation observed in clinical or test behavior may serve to identify or discriminate signifi-

cant classes of individuals and personality variables. Thus Hutt, Gibby, Milton, and Pottharst (10) and others suggested that the relative consistency or inconsistency of the Rorschach scores might differentiate, respectively, the "rigidity" of the abnormal subject from the "capacity to shift" or flexibility of the normal. Dörken (5) suggested modifications on a similar hypothesis depending on whether individual differences in variation were observed on an intra- or intertest basis involving projective techniques of different types.

The crucial question with regard to any proposals of this character is whether or not individual differences in variation are themselves generally stable and thus define true differences and not merely chance distinctions. Reviews of the literature on projective techniques reveal no effort to deal with this important question. The present study offered implications relating to the problem in terms of intercategory comparisons of individual differences in variation between high and low levels of movement production. The danger of using this method of observing individual differences in variation has already been noted. In the absence of any previous effort to deal with the problem of their stability in the projective techniques, however, the tentative results may be of considerable interest.

The results suggested that individual differences in variation were not significantly related from one area of performance to another, and that there was no basis for assuming that the differences were of general psychological import or that they defined a psychologically significant class of subjects or personality variables in any instance—except where the subject's performance was generally

consistent. The results thus offered no support among the movement indices for any proposals of the character suggested. At the same time, the results warned against making any diagnostic inferences on the basis of individual differences in variation without first checking on the stability of those differences.

The data did not preclude the possibility that individual differences in variation might be of a stable character *within* specified areas of performance as defined, for example, by each of the several categories of movement examined in the present study. It is not inconceivable that interpretive inferences might be warranted on such a basis. Confirmation of this possibility in the present experiment, however, would have required replication of the experiment with the same subjects and was beyond its scope. The possibility thus remains for further investigation.

The Movement Response in Relation to the Test or Type of Test

While the over-all results of the experiment suggested that interpretation of the movement responses given by ostensibly normal subjects may be justified for some few cases only, subsidiary analyses suggested that the number might be expected to vary depending on the test or type of test used. Of the two types of tests used in the experiment, the perceptual tests appeared to offer the greater promise for discriminating substantial numbers of such subjects. This implication was evident on the basis of the positive trends observed in the perceptual tests, as contrasted with the negative results among the verbal tests. It might be concluded that the perceptual tests somehow impose conditions on the subject's performance which are more conducive

to tapping the deeper rooted, more basic aspects of personality and behavior as related to the movement response and, accordingly, possess more in common as measures of movement response tendencies.

Of the three perceptual tests, the Rorschach and Behn-Rorschach appeared to enjoy greater mutual relationship in the area of movement production than any other pair. Their mutuality appeared to extend through every movement category, including sum movement—which, incidentally, appeared to be the most stable of the several indices in this relationship, whether observed on an intra- or intertest basis. It may be noted that Epstein, Nelson, and Tanofsky (7), using their extensive series of new inkblots, similarly found sum movement to be the most reliable of the movement indices. The results of this as well as of the present study might be considered to support those authorities who insist that important information may be overlooked in scoring only the human movement responses. The results would suggest, further, that if preference is to be given to any one index, it should perhaps be given to sum movement and not to human movement.

Considering the phi values obtained in the present study, however, in relation to correlation coefficients of comparable grade and in relation to the reliabilities of the two tests, there would appear to be some justification for regarding the relationship between the Rorschach and Behn-Rorschach to be approximately unity in all categories of movement when corrected for attenuation. (Incidental computations, in fact, confirmed this possibility.) There thus appeared to be no theoretical ground for preference between movement categories as far as these

two tests are concerned. It might be concluded that the two tests offer essentially common measures limited in their interrelationship in the several categories of movement only by their variously low reliabilities. While this finding encouraged acceptance of the tests as theoretically parallel forms for observing movement production, it would seem doubtful whether in their present state they can be reliably used as such in individual cases. In stressing this limitation, the results of the experiment essentially agreed with those reported by Eichler (6) and others for the two tests generally.

Similar but somewhat narrower conclusions appeared warranted concerning the Rorschach and Levy Movement Cards on the basis of their significant but low interrelationship in the *M* and *m* categories. It is notable that previous comparisons between these two tests have yielded no significantly positive results. Assessing their own investigations in this respect, King (12), and Allen, Ray, and Poole (1) suggested that the negative findings might have been due to the marked differences in the instructions for the two tests. The results of the present experiment would support such a conclusion and suggest that the two tests do have some mutual validity when less structured directions are substituted for the standard form prescribed for the Levy test.

The contrasting results observed for the two types of tests, perceptual and verbal, lead to some speculation concerning the possible causes. Most readily apparent is the fact that the contrast did not appear to be due to differences in reliability between the two types of tests. The results were, in fact, contrary to what might have been predicted on these grounds, since the verbal tests were on

the whole more reliable than the perceptual tests.

Apparently, the stability of the movement responses as observed in the perceptual tests is sufficiently great as to make itself evident (for some subjects) despite the low reliability of the tests. If so, then the development of more adequate tests of this character, possibly along the lines suggested by Barron (4) might greatly expand their discriminative value. This possibility suggested itself most strongly in connection with the *m* responses which, despite being least reliable, revealed significantly positive results not only with respect to performance in the perceptual tests, but also for the full battery. It is conceivable that a successful effort to raise the reliability of the *m* category could make it the most generally significant; some special study may well be due this category of movement.

As to the possibility of raising the reliability of any of the movement indices, it is interesting that the results obtained by Epstein, Nelson, and Tanofsky (7) discouraged the feasibility of substantially improving reliabilities in the ink-blot tests by lengthening the stimulus card series within practical limits. It might be added that if the results of the present study can be accepted as indicative, then it may not be possible to raise reliabilities enough in tests of this type to permit individual prediction in any case except with reference to specific populations.

Except for the *FM* and *m* categories in Sentence Construction 2, reliabilities in the verbal tests appeared high enough to suggest that the tests might very well measure internally consistent tendencies of some sort. Apparently, however, the conditions imposed on the subject's per-

formance by these tests were not conducive to tapping such deep rooted and stable tendencies as might—except for some few subjects—generalize broadly either among the verbal or from the verbal to the perceptual tests. Or, if basically stable functions were indeed sampled, then the functions must be assumed generally independent of each other and of any sampled by the perceptual techniques.

Dörken (5) suggested that different types of projective test materials or tasks might tap different levels of personality structure. He asserted on this basis that projective techniques might be expected to yield mutually consistent results only when (a) the tests are of the same or of a similar type and hence measure on the same level of personality or (b) when the different levels of personality tapped by different types of tests are themselves consistently organized—a condition he considered likely only for normal subjects.

The general lack of agreement between the verbal and perceptual tests in the present experiment might partially, perhaps, be understood on Dörken's (5) terms. Failure of the majority of the subjects to be consistent within *either* of the two types of tests, however, is not so easily explained. It might even be concluded from this result that, quite contrary to Dörken's (5) principles, the projective techniques may measure on different levels of personality regardless of whether the same or different types of stimulus materials and/or tasks are employed. Some investigators have, in fact, suggested that the projective techniques may sample different levels of personality structure even within the confines of a single test, depending on the subject. Macfarlane and Tuddenham (14) have noted this possibility as, perhaps, a major

problem inherent in all projective techniques and a key question in relation to general theory and practice.

Implications for General Theory and Practice

Among the basic assumptions commonly accepted as logical consequences of the relatively unstructured nature of the projective techniques, Macfarlane and Tuddenham noted the following: (a) "... a belief that the psychological determinants of each and every response are basic and general." (b) "... the belief that projective techniques tap the durable essence of personality equally in different individuals" (14, p. 34). As against these assumptions, the authors noted that nonprojective psychology has long recognized test behavior as partially determined by peripheral and transitory factors. On their own account, they cited unpublished materials suggesting that, while the protocols of some subjects dwelt more on the basic and enduring levels of personality, others dwelt more on superficial levels and sometimes revealed only the protective coverings. In confirming the hypothesis of consistency for some subjects but not for others, the results of the present experiment easily lent themselves to similar interpretation, suggesting that the movement responses in a given projective test might well relate to basic, enduring levels of personality for some subjects but only to peripheral and transitory aspects for others.

The results of the experiment thus joined in suggesting that one of the important consequences of using relatively unstructured techniques may be that the subject is free to determine, consciously or unconsciously, by his own immediate set the level of personality on which he responds and, thereby, to limit the rela-

tion between his performance and basic aspects of personality and behavior.

Such an interpretation, however, involves acceptance of a prior assumption: that the test invariably samples personality functions even when the subject's performance departs from a basic frame of reference. Macfarlane and Tuddenham have expressed it as an explicit principle in "The Assumption of Response Determination." That is, "... that every subject's responses are not the consequence of sheer accident but are determined by psychological attributes of that subject" (14, p. 33). This statement is comparable to that offered by Zubin and Young, namely "... that the test behavior itself is a small sample of life behavior and not a random inexplicable performance" (30, p. 2-5). Macfarlane and Tuddenham (14) stressed the principle as a statement of scientific determinism, which few psychologists would be likely to question. Zubin and Young (30) emphasized it as a postulate which demands proof.

It may be noted that the principle as expressed in both statements predicates the personal relevance of all classes of projective test performance on a hypothesis of nonchance behavior. Treatment of the data in the present study offered a test of the hypothesis and, thereby, of the principle itself, in relation to the movement responses. The results showed that only for a very small number of cases did the movement responses occur with such regularity as to minimize chance occurrence. It might be concluded accordingly that the movement responses might very well be "determined" by chance for some and even the great majority of subjects in certain groups. By the same token, it might seriously be questioned whether the non-

chance hypothesis and, with it, the assumption of response determination can be accepted, without qualification, as generally valid for all classes of projective test behavior.

One of the most important implications of the experiment was thus to stress the need for basic research on the projective techniques with a view toward ultimate verification or revision of their commonly accepted theoretical foundations. The results of the present study suggested that a safer premise for observing projective test behavior might be the proposition that the subject's responses *may* or *may not* be "determined" by chance. Pending reversal of this implication by further research, it might be concluded that the interpretation of the movement responses and, perhaps, other classes of projective test behavior, as well, may not be justified without establishing in the individual case that the subject's responses actually occur in accordance with nonchance principles. Measures of intra-individual consistency based on multiple or battery testing may thus prove to be indispensable for both experimental and clinical purposes.

SUMMARY

The purpose of the experiment was to examine the assumption that the movement response as observed in projective tests represents something basically stable in the individual's personality and behavior.

It was hypothesized that, if the assumption is tenable, then the individual's comparative level of movement production is consistent over a range of different test situations.

The hypothesis was examined on the basis of the percent human (*M*), animal (*FM*), inanimate (*m*), and sum movement

(*SM*) scores obtained by 82 ostensibly normal male military recruits, 18 to 25 years old, on a battery of three perceptual and three verbal tests: the Rorschach, Behn-Rorschach, Levy Movement Cards, two new Sentence Construction tests, and a modification of Severn's heteronym technique. The data in each category of movement were first examined for evidence of significant interest agreement on whether the subject scored consistently above or below the median on the full battery of tests. Supplementary analyses were made to detect trends and relationships among the three perceptual tests, the three verbal tests, and all possible pairs of tests. Finally, individual differences in variation were compared from one category of movement to another. Two additional hypotheses were examined in this connection: (a) that the degree of variation shown by the subject in one category of movement is consistent with that shown in other categories; (b) that all or a significant majority of subjects are consistent in some one or more categories of movement.

It was concluded from the results that:

1. The conception of movement in all three common categories—*M*, *FM*, and *m*—might well be related to deep rooted, stable characteristics in some but not for the great majority of the subjects.

2. The relationship appeared more likely to be found in tests of the perceptual than of the verbal type, but even on this basis it did not appear likely for all or even a majority of subjects.

3. If the conception of movement bears any significant relationship to psychological processes for more than a minority of subjects, then the relationship is largely specific to certain tests or types of tests and test performance is not generally predictive from one type of test

to another for any individual.

4. Individual differences in variation were not of a generally consistent character from one category of movement to another so as to suggest that they might represent something intrinsically stable in the individual except where the subject is generally consistent in level of movement production.

5. There was no indication in inter-category comparisons that stable individual differences might be generally defined on the basis of selectively consistent tendencies favoring some one or more of the several categories of movement.

The results of the study were considered in relation to others reported in the literature and were cautiously interpreted to suggest that:

1. The conception of movement might

well be relevant for some individuals or groups but not for others; future research might well be directed toward identifying these classes of subjects and providing a basis for discretionary interpretation in clinical practice.

2. Tests of the perceptual type appeared to offer greater promise than those of the verbal type for discriminating substantial numbers of such subjects.

3. The utility of individual differences in variation for diagnostic classification—if warranted, at all—is limited to specific areas of performance and cannot be assumed without establishing the stability of the observed differences.

4. The validity of certain basic assumptions underlying the projective techniques was questioned and their revision suggested.

REFERENCES

1. ALLEN, R. M., RAY, C. D., & POOLE, R. C. The Levy Movement Test: Suggestions for scoring and relationship to the Rorschach movement responses. *J. consult. Psychol.*, 1953, 17, 195-198.
2. ALTUS, W. D. Some correlates of the group Rorschach and the schizophrenia scale of the MMPI among two groups of "normal" college students. *J. consult. Psychol.*, 1948, 12, 375-378.
3. ALTUS, W. D., & THOMPSON, GRACE M. The Rorschach as a measure of intelligence. *J. consult. Psychol.*, 1949, 13, 341-347.
4. BARRON, S. Threshold for perception of human movement in inkblots and its personality correlates. *J. consult. Psychol.*, 1955, 19, 33-38.
5. DÖRKEN, H. JR. Projective tests and the consistency of the personality structure: A pilot study. *J. abnorm. soc. Psychol.*, 1953, 48, 525-531.
6. EICHLER, R. A comparison of the Rorschach and Behn-Rorschach inkblot tests. *J. consult. Psychol.*, 1951, 15, 185-189.
7. EPSTEIN, S., NELSON, JANE V., & TANOFKY, R. Responses to inkblots as measures of individual differences. *J. consult. Psychol.*, 1957, 21, 211-215.
8. HARROWER, MOLLY R., & STEINER, M. E. *Manual for psychodiagnostic inkblots (a series parallel to the Rorschach inkblots)*. Author.
9. HAYS, W., GELLERMAN, S., & SLOAN, W. A study of the relationship between the verb-adjective quotient and the Rorschach experience balance. *J. clin. Psychol.*, 1951, 7, 224-227.
10. HUTT, M. L., GIBBY, R. G., MILTON, E. O., & POTTHARST, K. The effect of varied experimental "sets" upon Rorschach test performance. *J. proj. Tech.*, 1950, 14, 181-187.
11. IVES, VIRGINIA, GRANT, MARGUERITE Q., & RANZONI, JANE H. The "neurotic" Rorschachs of normal adolescents. *J. genet. Psychol.*, 1953, 83, 31-61.
12. KING, G. C. Rorschach and Levy Movement responses: a research note. *J. clin. Psychol.*, 1955, 11, 193-195.
13. KLOPFER, B., & KELLEY, D. McG. *The Rorschach technique*. Yonkers: World Book Co., 1946.
14. MACFARLANE, JEAN W., & TUDDENHAM, R. T. Problems in the validation of projective techniques. In H. H. Anderson, & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 26-54.
15. MAYMAN, M. A. A comparative study of the Rorschach, Harrower and Behn-Eschenburg inkblot tests. Unpublished master's thesis, New York Univer., 1947.

16. MELTZOFF, J., & LEVINE, M. The relationship between motor and cognitive inhibition. *J. consult. Psychol.*, 1954, 18, 355-358.
17. MILLER, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.*, 1953, 48, 367-375.
18. MILTON, E. O. The influence of varied experimental sets upon certain Rorschach variables: The human movement variables. Unpublished doctoral thesis, University of Michigan, 1950.
19. RABIN, A. I., & SANDERSON, M. H. An experimental inquiry into some Rorschach procedures. *J. clin. Psychol.*, 1947, 3, 216-225.
20. RORSCHACH, H. H. *Psychodiagnostics*. Bern, Switzerland: H. Huber, 1942.
21. RUST, R. M. Some correlates of the movement response. *J. Pers.*, 1948, 16, 369-401.
22. SANFORD, F. H. Individual differences in the mode of verbal expression. Unpublished doctoral dissertation, Harvard College Library, 1941.
23. SEVERN, D. J. Heteronyms for diagnostic use. *J. consult. Psychol.*, 1949, 13, 144-145.
24. SINGER, J. L., MELTZOFF, J., & GOLDMAN, G. C. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, 16, 359-364.
25. THORNTON, G. H., & GUILFORD, J. P. The reliability and meaning of Erlebnistypus scores in the Rorschach test. *J. abnorm. soc. Psychol.*, 1936, 31, 324-330.
26. TUCKER, J. E. Rorschach human and other movement responses in relation to intelligence. *J. consult. Psychol.*, 1950, 14, 283-285.
27. WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
28. WITTENBORN, J. R. Statistical tests of certain Rorschach assumptions: The internal consistency of scoring categories. *J. consult. Psychol.*, 1950, 14, 1-19.
29. ZUBIN, J. Symposium on statistics for the clinician. *J. clin. Psychol.*, 1950, 6, 1-6.
30. ZUBIN, J., & YOUNG, KATHLEEN M. *Manual of projective and cognate techniques*. Madison: College Typing Co., 1948.
31. ZULLIGER, H. *Einführung in den Behn-Rorschach test*. Bern, Switzerland: H. Huber, 1946.

(Accepted for early publication April 12, 1958)

APPENDIX

SENTENCE CONSTRUCTION 2

Stimulus words in order of presentation

1. the	11. then	21. first	31. with
2. all	12. because	22. also	32. in
3. once	13. least	23. any	33. between
4. both	14. or	24. within	34. so
5. most	15. across	25. as	35. another
6. often	16. only	26. though	36. about
7. every	17. although	27. but	37. either
8. a	18. anything	28. behind	38. therefore
9. some	19. always	29. how	39. again
10. each	20. however	30. and	40. very

HETERONYMS⁴ AND MASKING WORDS IN ORDER OF PRESENTATION⁴

Set 1

1. develop	6. <i>conduct*</i>	11. deliver	16. <i>reprint</i>
2. <i>insert</i>	7. examine	12. <i>protest</i>	17. awaken
3. continue	8. <i>increase</i>	13. assemble	18. <i>commune*</i>
4. <i>recruit**</i>	9. accomplish	14. <i>desert</i>	19. resemble
5. discover	10. <i>escort*</i>	15. distinguish	20. <i>offset</i>

Set 2

1. museum	6. <i>suspect*</i>	11. attorney	16. <i>combine*</i>
2. <i>converse*</i>	7. physician	12. <i>convict</i>	17. mechanic
3. apartment	8. <i>extract*</i>	13. horizon	18. <i>discount</i>
4. <i>import</i>	9. cathedral	14. <i>insult*</i>	19. umbrella
5. potato	10. <i>progress</i>	15. piano	20. <i>consort</i>

Set 3

1. together	6. <i>buffet</i>	11. tomorrow	16. <i>conscript</i>
2. <i>perfume*</i>	7. unjustly	12. <i>present</i>	17. wherever
3. extremely	8. <i>survey</i>	13. already	18. <i>ferment*</i>
4. <i>permit</i>	9. whenever	14. <i>compress*</i>	19. precisely
5. beforehand	10. <i>exploit</i>	15. exactly	20. <i>compact</i>

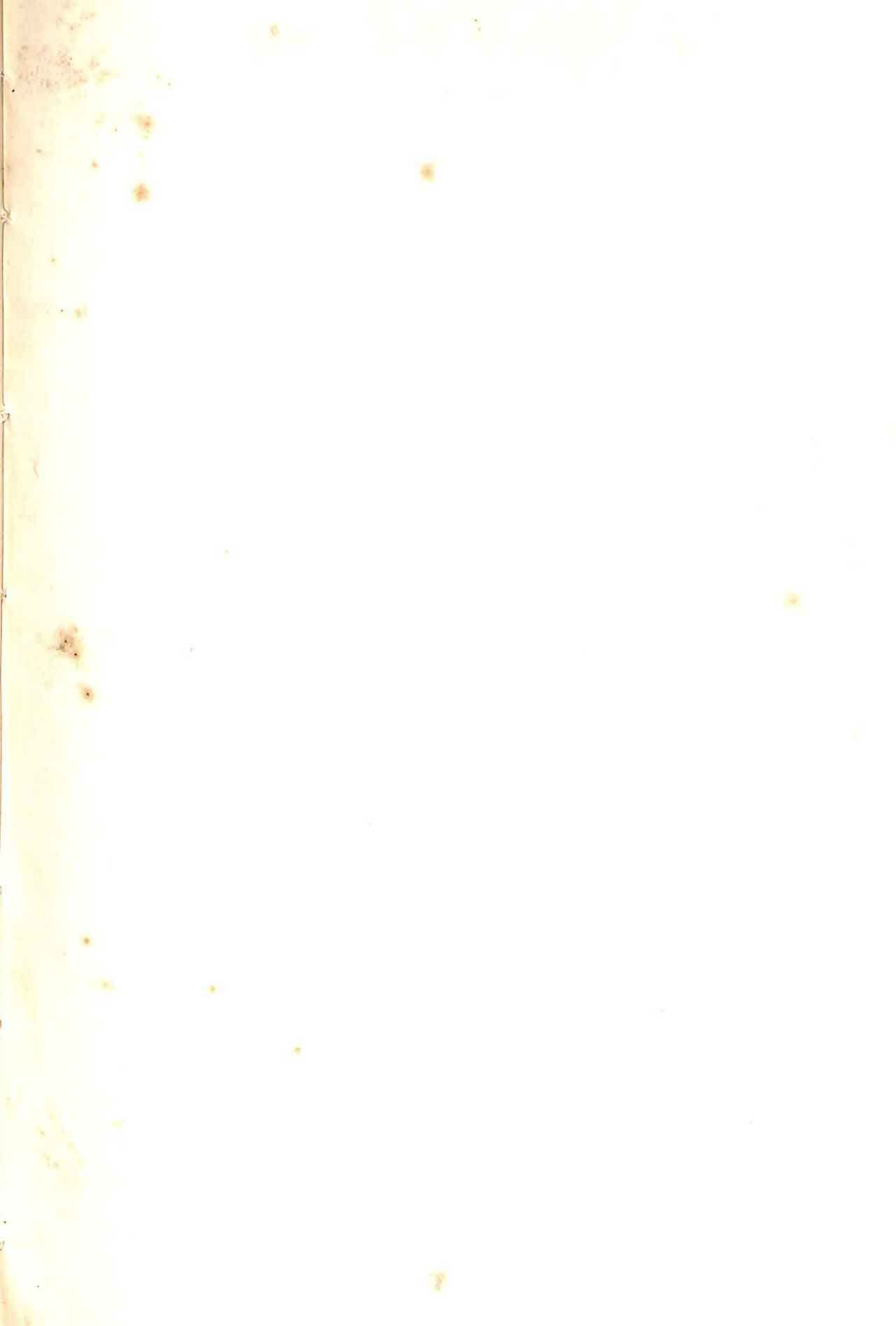
Set 4

1. romantic	6. <i>transfer</i>	11. unhappy	16. <i>console</i>
2. <i>convert</i>	7. dramatic	12. <i>produce</i>	17. abundant
3. familiar	8. <i>project*</i>	13. eternal	18. <i>recount</i>
4. <i>rebel</i>	9. tremendous	14. <i>abstract</i>	19. eleven
5. religious	10. <i>object</i>	15. peculiar	20. <i>digest*</i>

REPRODUCTION OF HETERONYM CARD 2, SET 1

w v n k g i y k g v m j n m o i x d b p l f j v t v
t m l f k c g d z w j l q d r h t a v x v x r z r y
r t s a j k h x f j i n s e r t g h i v w l a z q e
w d d p u v y r u l r g t h h v t c g q f v r f m z
z h w w g r f f b h f s y o g e v x e y n w z k e i
g w x y r p g h c i d y s k h l n v r w u x j k v i
v v i o z x y y p t w a f c x a q s q u j f i w s p

⁴ Italicized items: all except those asterisked are from Severn (23). Double asterisked item erroneously included; not a heteronym.



Psychological Monographs: General and Applied

Patterns of Mental Functioning Associated with Prejudice in Children^{1,2}

BERNARD KUTNER

Albert Einstein College of Medicine, Yeshiva University

IN HIS scholarly analysis of the determinants of prejudice, Allport (2) says: "Whatever our values may be, prejudice is a *fact of mental organization* and a *mode of mental functioning*." Among the psychological characteristics peculiar to prejudiced thinking he notes the following: fallacious overgeneralization, prejudgment, misjudgment, stereotyping, perseverative mental sets, categorical thinking, and a need for certainty (3). While some of these cognitional factors have been found to vary directly with a high degree of prejudice and/or ethnocentrism, neither the covariance of the two sets of variables nor the "prejudice syndrome" has been conclusively de-

lineated. This monograph examines the interrelationships of prejudiced and unprejudiced attitudes in seven-year-old children and their characteristic patterns of thought and problem solving. In so doing, we have deliberately ignored socio-cultural and other environmental factors which may bear on the development or functional utility of prejudice. For purposes of study, we have given central importance to the cognitive context out of which attitudes and behavior mature. The viewpoint adopted here is that both "structural" (i.e., cognitional) as well as "dynamic" (i.e., motivational and cultural) factors lie at the roots of many hostile attitudes. It is probable that cultural and motivational factors may profoundly influence the development of cognitional capacities and those in turn serve to channelize and direct the former factors (34).

However, it is perhaps a preliminary step to more dynamic analyses to ascertain the nature and functions of the cognitional factors that are related to ethnic prejudices. Similarly, we should like to determine whether special cognitional processes are evident in persons showing little tendency to ethnic prejudices.

SOME DEFINITIONS

Before proceeding to a discussion of the problems we intend to investigate, we shall first attempt to delimit the meanings of various terms that are used in the course of this report.

¹ This monograph is based upon a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Social Psychology in the Department of Social Relations, Harvard University, 1950.

² The author wishes to express his gratitude to Gordon W. Allport who guided the course of this study throughout. Appreciation is greatly felt to the school system of Brookline, Massachusetts, Ernest T. Caverley, Superintendent; James R. Hobson, Chief School Psychologist; Elgie Clucas, Principal of the Michael Driscoll School; and to the teachers and children of the second grade of that school during 1949-1950 for the splendid cooperation received during the investigation. For their generous assistance at different stages of this study, indebtedness is felt to Else Frenkel-Brunswik of the University of California, Robert A. Harris of Brooklyn College, Gardner Murphy of the Menninger Foundation, and Robert Sears of Stanford University. This study was carried out while the author was the recipient of a Sigmund Livingston Fellowship of the Anti-Defamation League of B'nai B'rith.

By *prejudice* we shall mean: a readiness to respond to an individual or a group in terms of a faulty, inflexible generalization, the net effect of which is to place that person or group at some disadvantage. Thus, a prejudice would be such a generalization and a prejudiced individual would be a person holding such generalizations. Also, in terms of this definition, ethnic prejudice refers to the fact that the generalizations concern an individual as a member of a particular ethnic group, or concern that ethnic group itself. Other varieties of prejudice mentioned in the report similarly refer to the generalizations held in regard to the particular referent or object of the generalizations.³

By *cognitive process* we shall mean: such functions as reasoning, thinking, judgment, and the solution of problems. We omit from this definition any reference to perception, since we do not deal with perception directly.

By *ethnocentrism* we shall mean: the readiness to respond to an individual, group, object, or symbol identified with the self in terms of faulty generalization; the net effect of the generalization being to place the individual, group, object, or symbol at some special advantage.

For convenience we shall refer to all groups, whether religious, racial, cultural, or national as "*ethnic groups*." Hence, the prejudices toward various groups found in the study are henceforth ethnic prejudices. Naturally, we do not mean to imply the unity of all varieties of prejudice. Indeed, we support the contention that different types of prejudice may have different dynamic histories. But in young children, the subjects of our study, we believe that we would not be far amiss in saying that anti-Negro or anti-Jewish attitudes do not have profoundly different sources for the majority.

STATEMENT OF THE PROBLEM

If prejudice may be thought of as a readiness to respond with a faulty generalization, one might attempt to assess whether or not faulty generalizations are to be found *generally* in the individual's thinking. In terms of this hypothesis, prejudiced social attitudes would simply be special cases of a general incapacity to reason logically. Consider the nature of

generalizing. The individual makes a group of observations of a number of different events; an effort is made to relate the events in some way; the individual stimuli are bound together by some common bond—a concept—and the generalization (an integration of discrete stimuli) is induced. This process, commonly termed concept formation, enables the individual to think and act in a more efficient fashion than would be the case if multitudes of unique observations of events were made that in no way could be considered as part of a class or species of events.

Considered as a set of inadequately formed generalizations, prejudices may then arise, in part, from a tendency to form inadequate concepts. We would therefore like to test the hypothesis following from these observations:

1. In situations requiring the formation of general concepts or generalizations, more prejudiced children tend to be deficient in forming adequate concepts; less prejudiced children tend to develop adequate concepts.

Thus, in solving inductive problems we shall be interested in the *process* of reasoning as well as in the outcome of the process: the attainment of the concept involved. We shall be particularly concerned with the "hypotheses" developed by each subject, how often they change, to what extent they persevere, how aids to reasoning given by *E* are utilized. The central aim shall be to determine whether significant differences in inductive reasoning exist between children who are more and less ethnically prejudiced.⁴

It has often been asserted that prejudiced persons do not "think straight": somehow their

³ We agree in this definition with the position taken by Vickery and Opler (56) that more emphasis should be laid on misjudgment in the definition of prejudice, hence the inclusion of a phrase concerning the faulty nature of these generalizations.

⁴ Alternately, we shall refer to those high in ethnic prejudice as prejudiced or more prejudiced. Those low in ethnic prejudice are referred to as unprejudiced or less prejudiced.

logical processes do not function adequately when the object of their prejudices comes under discussion. Hayakawa (20) has offered numerous instances of cases in which incorrect (i.e., biased) conclusions are drawn on the basis of evidence that may or may not be sufficient for drawing conclusions. Morgan and Morton (33) have shown how personal convictions may distort conclusions drawn from syllogisms. From a cognitive viewpoint, then, prejudices may arise and be supported by a failure of deductive reasoning. An individual has acquaintance with person A who is a member of group X. He observes this person's behavior and decides that since person A is behaving thus-and-so, and person A is a member of group X, then all other members (or even that members *other* than A) of group X also behave thus-and-so.

Woodworth (60) observed that formal logic is not akin to the reasoning processes of human beings. "Its principles and rules afford essential checks on the validity of an act of reasoning but they do not picture the actual process of reasoning" (60, p. 807). In considering the reasoning of children, Werner (58) and others have shown that syllogistic logic is farthest from their actual thinking mechanisms. It is not usual even for adults to draw conclusions based upon clearly presented premises. Occasionally, during a discussion of a scientific problem, for instance, enthymemal syllogisms may be employed. Thus, deductions in their normal setting follow distinctly unsyllogistic and nonformal principles.

In terms of our problem, perhaps prejudices arise in part from some inadequacies in the deductive processes; incorrect or invalid conclusions are drawn *generally* on the basis of evidence that may or may not be sufficient. From this we arrive at the following hypothesis:

2. In attempting to deduce conclusions from given premises, more prejudiced children tend to produce more numerous as well as more invalid conclusions than do less prejudiced children.

Again our interest will be divided between the *process* of deduction and the validity of the conclusions themselves. We shall attempt to learn whether the manner of presentation and the presence or lack of equivocality in the premises have a significant effect upon the conclusion reached. Here we come in contact with Frenkel-Brunswik's important prin-

ciple: intolerance of ambiguity. Do more prejudiced children draw inflexible conclusions on the strength of inadequate premises? Are they more adept at strictly formal logic? Do they resist ambiguous interpretations and make no allowance for alternate interpretations?

There remains one further aspect of the cognitive factors in prejudice. Tolman, in putting forward his revised theory of learning, describes what he has termed "cognitive maps" (55). In essence, the cognitive map is the "picture" the individual has drawn of the nature of the environment. It is the mental image of the "lay of the land" in the behavioral environment. Frenkel-Brunswik and her co-workers have found what amounts to a constriction of the cognitive maps of her ethnically prejudiced subjects (12, 14). Rokeach has termed this a "generalized mental rigidity" (44). In our definition of prejudice we have noted the relative inflexibility of the generalizations about people in which prejudiced persons engage. Perhaps this inflexibility is, as Rokeach has found, related to a general inflexibility of mind: a tendency for ideas or concepts to persevere even when they become inappropriate to the reality situation. In Rokeach's study, the longer and less appropriate method of solving a problem was adhered to in rigid fashion by the more prejudiced children.

Our hypothesis regarding mental rigidity is as follows:

3. Prejudiced children tend (a) to establish inefficient mental sets and (b) to employ inappropriate methods to break them; unprejudiced children tend (a) not to establish inefficient mental sets and (b) to employ appropriate methods to break them when they are established.

These are the problems this study sets out to investigate. Each hypothesis is taken up individually in the following sections and an attempt is made finally to bring the findings together into a partial explanation of the mechanisms of prejudice.

THE SCIENTIFIC SETTING

Human beings are generally capable of developing feelings of pleasure and displeasure, like and dislike, passion and

indifference regarding the widest possible array of stimuli in their environment. An individual may possess feelings of warmth and enjoyment toward many of his fellow men or he may despise and reject others. Normally, he comes into contact with comparatively few other human beings during his lifetime. It is remarkable then, that with the great array of potential objects of hostility available in the individual's environment, so many vent their hostile or rejective feelings upon persons or groups with whom they have had little or no contact. A person, a religious faith, a race—no segment is too large to be included as an object of hostility. How then, do such antipathies come about? What are their origins in the individual?

The attention of numerous investigators has been drawn to these problems. Some studies have examined the development and content of children's attitudes and have attempted to relate them to significant persons or events in the child's life. Lasker (26), in his early study of anecdotal reports of teachers, social workers, ministers, etc., analyzed the social influences bearing on the inculcation of attitudes in children. Horowitz (22), after a careful investigation of childhood attitudes in both rural and urban areas, concluded: "attitudes toward Negroes are now chiefly determined not by contact with Negroes, but by contact with the prevalent attitude toward Negroes." Zelig and Hendrickson (65) were similarly oriented when they studied the reasons advanced by children to account for their racial attitudes. Generally, home and school social influences were given as the "cause" of their attitudes. These and other studies are premised on the belief that the child "learns" to be prejudiced by adopting the prevailing attitude pattern from the context of his immediate social environment. The very young child allegedly develops specific dislikes but is too young to generalize them because he has not yet the conceptual capacity to comprehend groups. As his conceptual powers increase he forms ideas concerning human groups of people and if the groups he identifies are objects of hostility for his parents or his contemporaries, then he in turn will feel hostility toward them. This process might be termed a "culture contact" or "exposure" theory of prejudice. No consideration is given to the functions which such prejudices might serve. The child merely conforms to the

group norms and learns the prejudices in the same fashion that he learns of other characteristics of his world.

Although prejudices against various ethnic groups are widespread, there exists at the same time a theme of tolerance for diverse cultural, religious, racial, and national groups. Many parents make special efforts to "teach tolerance." Many children nevertheless seem to develop the very hostile attitudes whose occurrence their parents seek to prevent.

To explain why some children do not bend as the proverbial twig, the common childhood phenomenon of contrasuggestibility might be advanced as an explanation. But contrasuggestibility would not account for cases of prejudice where the parents place no stress upon the need for avoiding prejudice or where ethnic groups are never discussed in the child's presence. Neither could it account for the reverse situation in which the child fails to develop prejudices in a cultural setting imbued with numerous hatreds. Here the child may rebel against or be unaffected by the group norms. Such mechanisms as "suggestibility" or "exposure" will be recognized, not as explanations or causes of prejudice, but as conditions of prejudice dissemination. Certainly, some children parrot the bigots to whom they are "exposed" but the persistence of bigotry in the thinking of the growing individual suggests that prejudice may serve important functions in his adjustment.

Considerable work has been done to determine the ages at which social attitudes appear. Goodman (17) studied manifestations of racial attitudes in nursery school children. Radke, Trager, and Davis (41) traced the ways in which groups' attitudes were shaped between kindergarten and the fourth grade. Lasker (26), Criswell (11), Moreno (32), Horowitz and Horowitz (23), and Horowitz (22) carried out similar types of investigations with various age groups. Others such as Meltzer (30) and Zelig (63) studied time trends in group attitudes over several years.

Generally, these studies attempt to answer what may be described as "actuarial" questions: how prejudices appear at various age levels, how they change over time (most agree that they remain relatively fixed) and how variations in their manifestation, using alternate instruments, may be elicited. Little empirical work has been done to answer the question: What are prejudiced children like? And how did they get to be that way?

Some investigations into childhood attitudes have been concerned with the question of the relationship of prejudice to intelligence. Implicit in some of these studies has been the belief that tolerance is a function of increasing intelligence. In his comprehensive research on attitudes

toward the Negro, Horowitz (22) made a rough estimate of the relationship between hostility toward the Negro and intelligence. The attitudes of classes of "dull" and "bright" students were compared. No consistent differences were found. Zeligs and Hendrickson (65) computed the relationship between intelligence (measured by the Otis Group Test) and a measure of "generalized tolerance" administered to 178 sixth-grade children. They found $r = .31 + .05$. Minard reported (31) that 521 of 1,352 Iowa school children from the seventh through the twelfth grade had had intelligence tests already given them. By selecting 100 cases on a chance basis and computing a product-moment r of intelligence and scores on his B test,⁵ Minard found $r = .34 + .06$. Using the full sample and comparing scores on both the B and C tests, he found an increasing tolerance score significantly related to increasing intelligence. Horowitz (22), considering the differences between his own study and those of Zeligs and Hendrickson and Minard, decided in favor of the latter two studies in a notable display of scientific objectivity:

"In the light of the crude estimate of intelligence and the casual way in which the comparisons were made, it is suggested that Horowitz's data be eliminated from consideration in this context in favor of the tentative conclusion that there is evidence to support the hypothesis that brighter children tend to be relatively more tolerant than duller children of equivalent age" (p. 178).

Frenkel-Brunswik (12), refining the analysis of the intellectual component, reports that intelligence of 120 children aged 11 to 16 was only slightly correlated with ethnocentrism (prejudice) as measured by the California Attitude Scale I, but that it was found to be relevantly related to *rigidity in thinking* (p. 303). General intelligence as measured by IQ ratings appeared not to be as significant a covariable of ethnocentrism as was some special subfunction of intelligence designated as "rigidity" in the mental processes. The issue of intelligence as a covariate of prejudice remains in doubt, although, of the studies discussed, the California research was the most extensive. Because of its relevance to the present research we examine here, in some detail, the findings of the California Study.⁶

⁵ Minard computed three scores for his test. The A score was based upon the full test. The B score was based upon 71 items involving moral judgments. The C score was computed from responses to items involving personal preferences (i.e., as in race of marital partner).

⁶ The California Study referred to in this report comprises a series of investigations on personality and cognitive variables in the prejudiced personality, carried out at University of California at Berkeley (see 1, 12, 13, 14, 15, 44).

The concept of rigidity as used in the California Study may be understood as a general orientation of the mental functions to select and persevere at a given idea, response tendency, belief, or expectation about the environment. A cognitive perseveration is seen as part of a tendency for rigidity to generalize into all areas of the individual's life. Thus it was found that more ethnocentric children tended to conceive of sharply dichotomous sex roles (i.e., "girls should only learn things that are useful around the house"). Masculinity was associated with strength and femininity with weakness. Furthermore, the California Study isolated what was termed an "anti-weakness" attitude of the prejudiced child. That is, the child preferred to be identified only with the dominant ingroup and to approve of the isolation or removal of "weaker elements" in the population.

The fact that the prejudiced child was found in this study to look with favor on ideas that tend to segregate his world and to dichotomize his experience into distinctly opposed groups (i.e., good and bad, acceptable and unacceptable, etc.), gives plausibility to the further finding that such children are "intolerant of ambiguity." Frenkel-Brunswik has given central importance to this concept in several papers (12-15) pointing out that the very nature of prejudice hinges on the child's incapacity to attribute multiple potential characteristics both favorable and unfavorable, to outgroups. The less prejudiced person can admit the existence of both sides of a question, can avoid universal evaluative judgments of others, even at the sake of self-consistency. Frenkel-Brunswik points out that intolerance of ambiguity is not equivalent to emotional ambivalence although the two are related. Emotional ambivalence involves the simultaneous desire to accept and reject, to agree and to disagree, and to love and hate. Intolerance for ambiguity precludes uncertainty. It involves rejection of those aspects of the environment that cannot be apprehended, given the individual's values and attitudes. Ambivalence refers to an emotional disposition to entertain *conflicting* feelings concerning an object; intolerance of ambiguity refers to the failure of recognition and rejection of conflicting or alternate potential characteristics in objects. Ambivalence toward the family was found to be associated with intolerance of ambiguity. Emotional ambivalence and cognitive intolerance of ambiguity each may be viewed as consistent aspects of a particular type of personality organization. Investigators, searching for relationships between personality functions and social attitudes, have tended to account for attitude structure and

discussed personality, carried out at University of California at Berkeley (see 1, 12, 13, 14, 15, 44).

attitude change in terms of personality organization.

Indeed, the role of personality factors in shaping attitudes has often resolved itself into the question: What *type* of personality is likely to find what sorts of social events congenial or disagreeable? To answer this question, personality types are isolated and their social sentiments deduced. An attempt to establish an empirical typological psychology may be found in the work of E. R. Jaensch and his followers (25). Jaensch, employing perceptual and psychological tests as a measuring rod, defined two basic perceptual (i.e., personality) types: the "liberal" type and the "integrated" type. The former is generally liberal in the political as well as "synesthetic" in the perceptual spheres. His percepts are "unstable, loosened up, even dissolved" (25, p. 110). On the other hand, the "integrated" type is found to possess "firmness, consistency and regularity" (25, p. 111) in both the cognitive functions and social attitudes. Needless to say, Jaensch viewed the latter type as preferred and as being closer to reality because there is not so much free play in memorial as in perceptual tests. He posited that each "type" had a "unitary style" integrating the perceptual, cognitive, and social spheres.

More recently, a number of investigations have looked into the relationships between social attitudes and certain other variables, both motivational and cognitional (27). Newcomb related changes in social attitudes to functions of certain personality variables in a college community (36). Murray and his co-workers (35) have demonstrated the relationships that hold between personality functions and sociopolitical sentiments.

Smith (52) notes the close relationships of personal values and political attitudes. His problem interests us in regard to the present study since it serves to point up the differences between this approach and that taken by the numerous studies⁷ associating "need," "value," "personal relevance," etc., to attitudes. Smith reasoned that personal values have been shown to direct and determine cognitive experience. Hence, since attitudes are the functional representatives of perception in relation to cognized objects, then personal values may also determine attitudes.

The present study never reaches the stage of defining the personally relevant values or needs of the subjects, although

⁷ Current literature in perception theory is waxing heavy with studies aimed to clarify these relationships (5, 7, 8, 29, 40). Still others are waging the battle for a "non-separatist" view of cognitional and motivational processes (21).

an attempt is made in this direction. The question that we aim to answer, namely, "What are the cognitional correlates of prejudice?" remains, nevertheless, as a legitimate and separable problem. Let us then examine the results of that aspect of the broader California Study dealing with memory and perception in the ethnically prejudiced child.

Forty-two sixth-, seventh- and eighth-grade subjects who had already been administered the California Attitude Scale I⁸ and other personality and social belief tests were presented with a story (termed a "milieu characterization") involving the relationships of school children and their attitudes toward recent arrivals in the school. The newcomers are characterized briefly. One is Negro and one Jewish. The story relates the behavior of the "old" group of children toward the "newer" ones. Conflicts, acts of aggression and defense are included, ending with the new children attempting to defend themselves against the older ones. The task is to recall and reproduce the story. The findings in general support the author's contention that the more prejudiced children tend to distort the story to conform to their convictions about each of the ethnic groups. The more prejudiced children select the aggressive sequences for recall significantly more frequently than do the less prejudiced children. Another important difference in recall is the greater number of omissions of significant aspects of the story, not fitting the usual stereotypes held by the low-prejudice children over the high-prejudice children. The interpretation given is that the low scorers emphasize the "positive and nurturant aspects" of the story rather than the many cases of overt aggression. It is further found that more prejudiced children tend to become concentrated on the details and description of the people in the story or to wander off into fantasy using the story as a stimulus. Thus, Frenkel-Brunswick concludes: "It is in this manner that a rigid, segmentary, cautious approach goes with one that is disintegrated and chaotic, sometimes one and the same child manifesting both patterns in alternation or in all kinds of bizarre combinations. As do negativism and distortion in general, both these patterns help avoidance of uncertainty, one of them by fixation to, the other by tearing loose from the given realities" (p. 126). Being concerned with demonstrating the general characteristic of rigidity manifested in many ways and under many conditions, Frenkel-

⁸ Fifteen hundred subjects between 11 and 16 years of age were administered this scale.

Brunswik did not distinguish between perseveration on a given theme, assimilation to a stereotype, and adherence to a set of prepared attitudes—three types of rigidity that appear to characterize the results reported above. The central factor of rigidity in the thinking of the more ethnically prejudiced children appears to have been well established, however.

The perceptual experiments consisted largely of presenting the subjects with ambiguous stimuli. Unfortunately, since these studies are not completely reported upon, it is difficult to evaluate the reports. One experiment with 14 subjects employed a disk-shaped reversible figure-ground pattern and measured time to shift and spontaneous shifting. The results, however, were inconclusive. In a second experiment, a series of pictures were presented in which the likeness of a dog gradually emerged as the image of a cat. The more prejudiced children perseverated, holding the image of the dog over a greater number of steps in the sequence than did the less prejudiced children. There was also a greater degree of haphazard guessing in the center of the series by the more prejudiced children. Additional studies were reported in which hues and numbers were progressively changed and similar results were obtained. Particularly interesting is the fact that the prejudiced child not only tended to cling to a given percept once it had been made but also devised means for ignoring or obliterating the incongruities of the changeable (i.e., fickle) stimuli before him.

In an extension of his study of "generalized rigidity" in ethnocentric adults, Rokeach (44) gave ethnocentric children two problem-solving situations involving the establishment of a mental set toward a particular solution. He demonstrated that the more ethnocentric child tended to adhere to the less appropriate solution when an alternate, more appropriate, solution was possible. The less ethnocentric child showed less rigidity by more frequently selecting the shorter solution which Rokeach defined as more "adaptive."

Several general conclusions may be drawn from the studies noted above. First, the prejudiced child appears to be cognitively more rigid than the unprejudiced child. He tends to select for recall and recognition only certain aspects of the stimuli bombarding him: they stabilize his phenomenal world into relatively fixed and dichotomized categories. Categorization assists the child to deal with the discrepant and vaguer aspects of his

environment that, objectively, do not fit the categories. He both avoids and distorts reality to maintain the necessary categorical views. When forced to face ambiguities, the prejudiced child tends to assimilate his percepts to one or the other of the alternatives and to persevere at the fixated alternative. Indefiniteness is apparently anathema to such children. Moreover, even when the alternative selected is inappropriate or maladaptive and the more appropriate and adaptive alternative is presented the individual fails to utilize it because of the tenacity of the original fixed set.

It must be pointed out that the relationships referred to above are by no means positively established as invariant aspects of the makeup of the prejudiced individual. Indeed, the patterns empirically established by the work of the California group have large variances as indicated by selected cases given intensive personality study. Although the majority of the findings mark off the groups from one another in fairly consistent manner, there are notable individual differences requiring dynamic analysis to be understood.

METHODOLOGICAL CONSIDERATIONS AND PROCEDURES

Plan of the Study

This investigation compares the performances of relatively prejudiced children and those of children who are relatively unprejudiced, on a group of tests, interviews, and ratings concerned with the functioning of the higher mental processes. The tests themselves are described in detail in the chapters that follow. Here we should like to discuss some of the general problems encountered in testing attitudes in young children. Also we should like to consider the subjects of

the study and the general procedures employed.

The first necessary task of the study was to set up some measure of the presence or absence of ethnic prejudices. We assumed that any test of such a variable, especially as it concerns young children, would not be likely to tap all or even most of the ways in which ethnic prejudices might be expressed. Pencil-and-paper tests could not be used since the subjects could not read or write sufficiently well. Some form of projective test suggested itself for convenient use in a situation defined for the child as a "game." Two tests were decided upon. The first is an information test, requiring the child to tell what he knows about various ethnic groups. We hoped with this instrument to discover whether knowledge of a group was necessary for a child to express definite attitudes toward it. Additionally we anticipated that some expression of attitude would be obtained spontaneously. The second test presents a group of statements attributed to someone else, who places various ethnic groups in strongly prejudicial or ethnocentric terms. We hoped here to obtain agreement or disagreement with the statement together with spontaneous reasons the child might advance for holding these views. The first test produced inadequate results and was discarded while the second test proved valuable and is the basis for our final groupings.

We decided that the subjects would be scored for prejudice on a roughly linear four-point scale. We followed this procedure for two reasons: (a) we wished to score each child in such a way that the more prejudiced group would be clearly different from the less prejudiced group. A four-point scale would first differentiate two extreme groups. The intermediary groups could then receive careful individual analysis before being categorized as either prejudiced or unprejudiced. (b) From a four-point scale we could proceed to classify each child into either a more prejudiced or a less prejudiced group. Our desire was to define only two groups: those high and those low in ethnic prejudices.

All children who were finally classified into either of the two categories at the low-prejudice end of the scale were called unprejudiced and formed Group A, while all children falling into either of the high-prejudice categories were classified as prejudiced and formed Group B.

It may be of some value to point out the differences in classification of more and less prejudiced children between the California Study and the present one. To test their hypotheses on cognitive and personality variables, the extreme quartiles of the scores on the California Attitude Scale I were selected. Of the individuals in these extremes, 120, or 8% of the

total of 1,500 subjects who were originally tested, were selected for intensive study. Hence, both the more prejudiced and less prejudiced groups contain the extreme 4% of the total at each end of the scale of prejudice. On the other hand, we administered all tests to all subjects, deriving a final two-point scale. Assuming that our test is valid and our categories reliable, children whose prejudice scores would lie at the medial part of the four-point scale would attenuate the findings no matter in which group they happened to be placed finally. If there are distinct "types" of prejudiced and unprejudiced persons, the results would not show up so clearly as they might if we had taken the extreme quartiles for special testing. We have attempted to overcome some of the difficulties inherent in our classification by defining the more prejudiced group in such a way that only those showing spontaneous and indisputable evidences of group antagonisms were classified in the B Group. The latter, consequently, may be considered to contain "pure" cases of prejudice. The A Group on the other hand not only contains the cases clearly classified as less prejudiced (at one extreme of the four-point scale) but also contains a number of doubtful cases whose classification was decided on only after the final rationale for the two groups was worked out. These considerations should be borne in mind in evaluating the results of this study. While our categories might imply that we are convinced of the existence of "types" of persons based on personality variables of which prejudice is one component, no such implication is intended. However, the nature of the study is such that conclusions may be drawn on this point even though final proof of any such dichotomy remains for further research. A paper by Frenkel-Brunswick (14) is aimed in this direction.

General Procedure

We first established contact with the School Department of the town of Brookline, Massachusetts (a large residential community lying adjacent to and south of Boston) and were put in touch with the Michael Day Driscoll School, one of eight public elementary schools. Arrangements were made to carry out the study with both second-grade classes as subjects. In conjunction with the principal and the two teachers involved, plans were made to collect the data over a five-month period. The *E* was introduced to the children as an assistant who would be helping the teachers during the year by trying out some new "games" in each of which they would have an opportunity to participate. During the first five weeks of the school year, the *E* acquainted himself with the children, permitting them an opportunity to

recognize him as a daily visitor. The *E* took his lunch with the second grade and went to recess with them each day, ensuring that he would be recognized as a regular member of the staff.

At a prescribed time, just prior to the start of the testing sessions, each class in turn participated in a discussion led by the *E* on the subject of "Peoples of the World," a topic chosen in connection with the social studies curriculum of the second grade. The discussion lasted about 25 minutes and about 80% of the children of each class participated by telling what they knew of the customs of various peoples. The discussion was intended to give the *E* a point of departure for the administration of the attitude tests which followed shortly.

When the testing sessions began the procedure adhered to was as follows: The *E* entered each class on alternate days at about the same time in the morning. A child was selected (all were taken in alphabetical order of the surname) and left with the *E* for his office. At the conclusion of a session, the child would be asked to bring to the office the child whose turn it was next. Sessions were held each day for about three hours, the *E* interviewing between five and nine children per day.

The problem of rapport, so frequently encountered in research or testing where the *E* has only limited contact with the child, was not a factor in this study. By the time the *E* began testing, he had been accepted as a normal part of the class structure. The start of the testing sessions found the children vying with one another for a chance to "go to the office" with the *E* for "games." The children soon learned that they were to go alphabetically, though some pleaded with the *E* to take them ahead of another whose turn fell earlier. This unique situation would seem to provide fertile ground for obtaining rich material, and the *E* was not disappointed in this regard. The subjects were natural and uninhibited, some overly exuberant at the chance to play "games" in the middle of the day. Most of the children attended to the tasks presented them and participated thoroughly. There were few exceptions to this rule. As a consequence few records had to be discarded for lack of responsiveness to the situations. In general, the research situation was unusually good.

The data were collected between October 10, 1949, and February 7, 1950. With a few exceptions each subject was seen on four different occasions. In certain cases where the data were not clear because of an interruption in the original testing session or some other extrinsic cause, a repeat session was held if the nature of the test permitted. A total of 245 testing sessions was held averaging 20 minutes per session. All testing was done individually.

Subjects

A total of 65 subjects participated. Complete sets of data are available for 57 subjects and partial data for 3 subjects who left school during the year. The remaining 5 subjects took the attitude test but none of the remaining tests, by reason of leaving school in two cases and because of indecisive responses on the attitude tests in the other three. Hence, the data reported upon in this study are drawn from the responses of 60 subjects.

There were 34 girls (57%) and 26 boys (43%) comprising the total group. Their average age at the start of the study was 6 years 11 months. The mean age of the boys was 6 years 10 months; that of the girls was 7 years 0 months, the differences being insignificant.

An analysis of the occupations of our subjects' fathers reveals (Table 1) that more than half either owned their businesses, retail and wholesale firms, or had supervisory or managerial positions in some business. Another group of 17 fathers were in some profession. Of this number, 5 were physicians and 5 attorneys. There were 2 engineers, a dentist, a pharmacist, an insurance broker, an accountant, a sports editor, and a teacher. A group of 9 parents were classified as "skilled." Of these, 7 were salesmen, 1 was a machinist, and 1 a toolmaker. In addition, there was one unskilled worker, a building superintendent, and there was one working mother who owned her own small business, the father being dead.

From a socioeconomic point of view, the majority of our subjects' families would fall into the middle to upper-middle class group. Most owned their own homes and had cars of recent model. Moreover, most of the homes were embellished with expensive furnishings, television sets, and other marks of middle-class status. The school itself, which drew its population from the surrounding neighborhood, is located in a suburban residential section of the town, the majority of the homes being privately owned. Rentals for the area were relatively high.

TABLE 1
OCCUPATIONAL GROUPING OF FATHERS
OF 60 SUBJECTS

Occupational category	N	%
Professional	17	28
Business-managerial	32	53
Skilled	9	15
Unskilled	1	2
Other	1	2
Total	60	100

From the point of view of religious preference, there were 3 Protestant children, 9 Catholic, and 48 Jewish children, representing respectively 5%, 15%, and 80% of the total. A word concerning this distribution is in order. Since this study concerns itself with the *general* associations between prejudices and certain cognitive functions, we are not particularly concerned over the ethnic make-up of our sample. Membership in a particular ethnic group should have no direct bearing on our problem. Frenkel-Brunswik et al. (1, 13) have found that the particular background of the child, including ethnic background and variety of familial structure, has a striking relationship with the child's personality organization and attitude structure. Nevertheless, we may still ask: Are there patterns of mental organization and functions that are related to ethnic prejudices and common to *all* ethnically prejudiced persons? If we find that such patterns of mental functioning do exist, we should feel more strongly that deep-seated prejudices may have some common base in mental structure irrespective of their specific cultural origins.

Why, it may be asked, are children the subject of this study and why are these children seven-year-olds? The California Study of ethnic prejudices studied children between the ages of 11 and 16 years of age. Split-half reliability of their responses on the scale testing ethnocentrism ranges from .82 to .90, "indicating that ethnic prejudice is a consistently and firmly established pattern not later than at the earliest age levels studied" (12, p. 296). We felt that, in order to study the earliest manifestations of ethnic prejudices, we should tap them when they are not yet consistent and rationalized. Similarly, we wanted to know whether special cognitive functions could be demonstrated as an early childhood concomitant of prejudice. Knowing the status of ethnic prejudices at various levels of development, we may assess the consistency of any relationship between prejudices and other variables at each stage. A question that begs for research is: Can we determine which variables are most salient in predisposing to prejudice at any age, and are these variables to be found also in the older, already prejudiced child? Are there diagnostic patterns to be found in childhood that will be predictive of prejudices? Our research is concerned with a necessary first step: defining the relationship of cognitive functions to ethnic prejudices but in *young children* rather than children approaching or in adolescence.

Another reason for selecting seven-year-olds is a finding of Allport and Kramer (4) that adults recall the origin of their prejudices as dating from the ages 6 to 16. The California Study touched the upper end of this range, while the

present investigation concentrates on the lower level. Together with the studies of Meltzer (30), Horowitz (22) and Zelig (64), etc., we are gradually finding confirmation of the Allport-Kramer finding. This study hopes to add further to the data on this point. Finally, we felt that since prejudices are undoubtedly only recent acquisitions in seven-year-olds, clues to processes and techniques aimed at reduction of—or prevention in the case of other as yet unprejudiced children—ethnic prejudices might be found at the earliest point at which prejudices manifest themselves.

ESTABLISHING THE PREJUDICED AND UNPREJUDICED GROUPS

Our first task was to construct a test that would measure the relative presence of prejudice. Under the circumstances in which the study was carried out, in the setting of a public school, it was felt best to rely on the responses to verbal material rather than employ free doll play or miniature life toys. We turn here to an examination of the tests and an analysis of the results.

Our measure of prejudice is an Ethnic Attitude Test, a structured projective questionnaire dealing with attitudes toward various ethnic groups. We decided to deal with only 7 major ethnic groups in constructing our test. Other items were added to make a total of 10 (Appendix A). After pretesting a number of different methods of presenting the questions on a group of eight-year-old children, we constructed a test consisting of statements that are highly prejudicial. Consistent with our definition, we felt that a prejudiced child would tend to agree with statements involving anti-ethnic group prejudices. On the other hand we felt that an unprejudiced child would tend to disagree with such statements. More prejudiced children would be likely to enlarge upon a prejudiced statement while less prejudiced children would be more likely to enhance their denial of the statement with facts or atti-

tudes indicating the acceptance or tolerance of the accused groups. Similarly, we felt that, together with the rejection of outgroups, more prejudiced children would reject statements antagonistic to their own group and espouse its virtues instead.

Only six of the seven ethnic groups dealt with in the questionnaire were employed to derive a final estimate of prejudice. These groups are Mexican, Catholic, Jewish, American, colored (Negro), and white. German, Italian, and Russian people were excluded from the criterion groups of the EAT because of the comparatively few subjects who could define these groups. Christian (Gentile) was also excluded due to the number of incorrect identifications (many of the children did not distinguish Christian or Gentile from Catholic). Chinese were excluded since a large proportion of the children had been stimulated by pictures and stories of Chinese in native costume in their class. It was felt that the remaining six groups would give a clearer estimate of prejudices. With the exception of the Mexican group, the children had had personal contact with members of each of the other groups. We decided to retain the Mexican group since we wished to determine whether prejudices found in attitudes toward the five other ethnic groups would also be present in regard to a group with whose members none of the subjects had ever had any personal contact.

The results of the EAT are presented below in two ways: (a) An analysis is made of the various responses according to whether or not the subject agreed or disagreed with the statement; and (b) An analysis is made of the classification of the subjects into "more prejudiced" and "less prejudiced" categories. First, we turn to an analysis of the results of the EAT on the agree-disagree dimension.

The responses are classified into 11 categories.

These are:

Category

(A) *Unconditional agreement*, the individual categorically agrees with the prejudiced statement.

(B) *Conditional or partial agreement*, the individual agrees with only some of the rejecting aspects of the statement or qualifies his agreement in some way.

(C) *Enhanced agreement*, the individual adds further rejecting remarks to those presented in the statement.

(D) *Partial agreement, partial disagreement*, the direction is equivocal.

(E) *Unconditional disagreement*, the individual categorically denies the prejudicial remarks made about the group.

(F) *Conditional or partial disagreement*, the individual does not agree with the statement but challenges a part of it or qualifies his disagreement.

(G) *Enhanced disagreement*, the individual denies the statement and adds to his denial by attributing characteristics to the group opposed to the statement or refuting it.

(H) *Disagreement on moral grounds*, the individual implicitly disagrees with the statement by saying in effect, "You must not say such things about these people."

(I) *Equalitarian disagreement*, the individual implicitly disagrees with the statement by remarking, for example, "They're as good as us" or "We're the same as they are."

(J) Don't know.

(K) Miscellaneous, in general irrelevant remarks.

Making use of these categories, Table 2 presents the results of the EAT. Of particular interest to this study is the fact that only 58 or 16.8% of the responses are clearly prejudicial in nature. However, since responses to statements on American and white people reflect a rejection of patently ethnocentric sentiments rather than a "rejection" of ethnic groups, we may eliminate "American" and "white" agreements to their respective statements and recompute the percentage of prejudiced responses. This re-computation reveals that there are 41 or 12% prejudiced responses.

If we combine columns, E, F, G, and I (categories of clear-cut disagreement with

TABLE 2
ANALYSIS OF RESPONSES ON EAT TOWARD SIX ETHNIC GROUPS

Ethnic group	Categories											Total
	Agree			Disagree								
	A	B	C	D	E	F	G	H	I	J	K	
Mexican N %	5 8.6	4 6.9	3 5.2	1 1.7	6 10.3	14 24.1	4 6.9	0 0	7 12.1	8 13.8	6 10.3	58
Catholic N %	2 3.4	4 6.8	1 1.7	3 5.1	4 6.8	6 10.2	5 8.5	6 10.2	18 30.5	5 8.5	5 8.5	59
Jewish N %	3 5.3	0 0.0	2 3.2	2 3.8	8 14.0	9 15.8	18 31.6	5 8.8	6 10.5	3 8.1	1 1.8	57
Negro N %	6 10.0	9 15.0	2 3.3	3 5.0	5 8.3	16 26.7	13 21.7	0 0.0	3 8.0	2 3.3	1 1.7	60
Subtotal	16	17	8	9	23	45	40	11	34	18	13	234
White N %	1 1.8	0 0.0	3 5.4	2 3.6	4 7.1	2 3.6	12 21.4	1 1.8	18 32.1	1 1.8	12 21.4	56
American N %	4 7.1	3 5.4	6 10.7	4 7.1	6 10.7	1 1.8	10 17.9	1 1.8	9 16.1	6 10.7	6 10.7	56
Totals N %	21 6.1	20 5.8	17 4.9	15 4.3	33 9.5	48 13.9	62 17.9	13 3.8	61 17.6	25 7.2	31 9.0	346

the statements presented to the subjects) we find that there are 142 responses, or 58% of the total, again eliminating the responses to "white" and "American." Of the 41 prejudiced statements, 12 are against Mexicans, 7 against Catholics, 5 against Jews, and 17 against Negroes. Divided on the basis of a "generally accepting" or "generally rejecting" attitude and eliminating questionable responses (especially some of the responses in categories D and H) we find (Table 3) that in order of "acceptance" are Jewish, Catholic, white, Negro, Mexican, and American. In order of "rejection" are Negro, American, Mexican, Catholic, Jewish, and, white.

It is interesting to note in Table 3 that Mexicans are rejected by 21% of the subjects even though no child had had any personal knowledge or acquaintance

with any Mexicans. Of the five subjects rejecting Jews, two are Jewish and three Catholic. In general there are about four times as many accepting attitudes as there are rejecting attitudes.

If again we regard only the first four ethnic groups in Table 3, we get a clearer picture of the pattern of acceptances and rejections. To bring this pattern into sharper focus, we present the results of accepting and rejecting responses for the Mexican, Catholic, Jewish, and Negro groups (Table 4).

The ratio of accepting to rejecting attitudes stands at about 3.7/1. Jews are accepted more and rejected less than either the Mexican or Negro group. p is less than .02 for the differences between the Jewish-Negro frequencies and is almost at the .05 level of probability for the differences between the Jewish-Mexi-

TABLE 3
ACCEPTING AND REJECTING RESPONSES BY SUBJECTS^a ON THE EAT

Ethnic group	Accepting	Rejecting	Don't Know	Other	Total
Mexican N %	31 54.4	12 21.1	8 14.0	6 10.5	57
Catholic N %	39 69.6	7 12.5	5 8.9	5 8.9	56
Jewish N %	46 83.6	6 10.9	3 5.5	1 1.8	55
Negro N %	37 64.9	17 29.8	2 3.5	1 1.8	57
White N %	36 67.9	4 7.5	1 1.9	12 22.6	53
American N %	26 50.0	14 26.9	6 11.5	6 11.5	52
Totals N %	215 65.2	59 17.9	25 7.6	31 9.4	330

^a Since there are typically only one or two responses per subject, this table represents the number of subjects responding to each ethnic group on the dimension of agreement and disagreement with the statement.

TABLE 4
ACCEPTING AND REJECTING RESPONSES TOWARD FOUR ETHNIC GROUPS

Ethnic group	Accepting	Rejecting	Don't Know	Other	Total N
Mexican N %	31 54.4	12 21.1	8 14.0	6 10.5	57
Catholic N %	39 69.6	7 12.5	5 8.9	5 8.9	56
Jewish N %	46 83.6	5 9.1	3 5.5	1 1.8	55
Negro N %	37 64.9	17 29.8	2 3.5	1 1.8	57
Totals N %	153 68.0	41 18.2	18 8.0	13 5.8	225

can frequencies.⁹ The Jewish and Catholic groups are not significantly different in terms of the number of acceptances and rejections by our subjects. The Catholic group's acceptance-rejection ratio is not significantly different from any of the remaining three, although p is just short of the .05 level for the Catholic-Negro pairing. These findings indicate that a hierarchy of acceptances and rejections exists with the Jewish group heading the acceptance order and Negroes leading the rejection list. Hence, the Jewish group is much less frequently rejected than are other groups, except for the Catholic group. This is to be expected since the major identification group is usually placed highest on most tests of attitude. The fact that nine of our subjects are Catholic increases the frequency of Catholic acceptances and concomitantly probably decreases the number of Catholic rejections.

Returning to Table 3, we find that the "white" group is rejected by only four subjects, but more than one-fifth of the responses are classified in the "other" column. One fourth of the responses to American are "rejecting." Less than half of the responses are "pro-American." These latter two results are somewhat surprising in view of the fact that all of our subjects are both white and American. It may be that the large number of irrelevant responses to the statement on "whites" reflects an indecision on the part of some of the children who, recognizing themselves as white, cannot agree that they are not as good as others in some way. It may also indicate that some of the children are confused that others should cast them in negative light. The large number of rejections of the statement concerning Americans seems to vio-

late the principle that one's own group is given highest status on attitude tests. It should be emphasized that disagreement with the strongly "pro-American" statement does not in itself constitute the "rejection" of Americans. Rather, it would seem to indicate a tendency by some children to avoid the idealization of the ingroup. A consideration of Column I of Table 2 would seem to confirm this view. It will be noted that fully 61 or 18% of the 346 responses are classified as "equalitarian" disagreement. The subject says in effect (if not in actuality) that "I'm not any better than or different from other people" and "There are no 'better' and 'worse' people, we're all alike." Hence a considerable number of our subjects would appear to be anti-ethnocentric.

One further problem remains: how consistent are the subjects in their acceptances and rejections of the various groups discussed on the EAT? The answer to this problem will lead us directly to our problem of selecting our two major groups since it will tell us *how many* prejudices an individual subject displays and *how strongly* they are held. In Table 5, we present the degree of individual consistency on attitudes toward four ethnic groups.

TABLE 5
DEGREE OF INDIVIDUAL CONSISTENCY OF
ATTITUDE TOWARD FOUR ETHNIC
GROUPS

Degree acceptance	N Cases	%
a. Accept 4	15	
b. Accept 3	12	
c. Accept 2	8	
Subtotal	35	58
d. Accept 3		
Reject 1	7	
e. Accept 2		
Reject 2	6	
f. Accept 1		
Reject 1	5	
Subtotal	18	30
g. Reject 2	3	
h. Reject 3	4	
Subtotal	7	12
Totals	60	100

⁹Chi square for the former pair is 6.22, and for the latter it is 3.57.

The combination of groups a, b, and c yields 35 subjects or 58% of the cases in which no prejudices are evident. Another group of 18% (or 30% of the cases) accept at least 1 group and reject no more than 2 groups; but, more important for the present study, this group shows varying degrees of inconsistency of attitude—they each have at least one rejection as well as at least one acceptance of an ethnic group. The third group is consistently rejecting in attitude. No individual rejected all four ethnic groups. There are fewer than four acceptances and rejections in some groups due to either lack of knowledge about an ethnic group or to the presence of irrelevant responses. Hence 70% of the subjects show marked consistency of attitude, either accepting or rejecting. Consequently, we can now establish that the 41 rejecting responses were made by 25 subjects, while 35 subjects made no rejecting responses whatsoever. Hence, using the EAT as our primary measure of prejudice, we can proceed to an analysis of the responses of the 25 subjects responding with at least one ethnic group rejection.

Our major task now remains: to select that group of children we are to call "more prejudiced" and another group that we are to call "less prejudiced." We have seen that there is a marked consistency in the direction of the attitudes of a large proportion of our subjects. Hence, we might select our two groups from the extremes of the distribution of responses to the EAT. Those accepting 4 or 3 groups would constitute our less prejudiced and those with a preponderance of rejections would make up our more prejudiced group. This technique has the added advantage of being highly objective. We chose, however, a different technique, somewhat less objective, that we nevertheless feel is at least equally valid. After reading through the responses of all the subjects a set of criteria was drawn up in accordance with our definition of prejudice. These criteria form a four-point scale from most to least prejudice. We do not assume equal intervals between the points, nor is this a Guttman-type scale. We believe, however, that it fulfills the requirements of

a continuum: the quality and quantity of the major variable dealt with in the scale increases in relatively systematic fashion from point to point. Such a scale has one major advantage, crucial to our study, over a purely objective method enumerating the quantity of ethnic prejudices: it takes into account the *intensity* of the attitude as judged by the *E*. Hence, one case of ethnic prejudice exhibiting a high degree of hostility against a single group may be rated as more prejudiced than another case in which mildly held prejudices against two ethnic groups are found. The criteria for the fourfold classification follow:

Group 1. *Very Unprejudiced*

- a. Denies unwarranted assertions of virtues of own group.
- b. Denies or enhances denial of negative traits ascribed to ethnic group.
- c. Gives equalitarian response to ethnic group comparisons.
- d. No evidence of spontaneous rejection of an ethnic group.
- e. Does not place ethnic group or member at a disadvantage or deny his virtues when he knows they exist.

Group 2. *Somewhat Unprejudiced*

- a. Does not spontaneously attribute virtues to ethnic groups.
- b. Does not deny attributed negative traits to ethnic groups.
- c. Engages in moralistic admonitions (you must be good to everyone; you mustn't say such things).
- d. Indirect or mild rejection of an ethnic group or individual who is identified by his group membership.
- e. Rejection on aesthetic grounds (I don't like the color of him).
- f. Moderate ethnocentrism.

Group 3. *Somewhat Prejudiced*

- a. Agrees with statements attributing negative traits to ethnic groups.
- b. Tends to generalize a dislike or rejection from individual to group.
- c. Multiple hostile attitudes toward ethnic groups.
- d. Strong display of ethnocentrism.
- e. Spontaneous enhancements of attributed negative traits.

Group 4. *Prejudiced*

a. Spontaneous expression of hostility against ethnic groups aside from statement.

b. Enhanced spontaneous additions to attributed negative traits with rationalizations.

c. Multiple hostile attitudes, intensely held.

d. Rejects or denies existence of any positive characteristics of ethnic groups.

e. Attribution of negatively valued traits (bad, smell, steal, fight, etc.)

In order to qualify for classification in a particular group, the child's responses must fall within at least one of the operational criteria given for each group. Hence, the appearance of only one of the items in each category is sufficient to warrant its classification in that category.

Two subjects could not be classified for various reasons and were consequently dropped from the list. In order to establish the reliability of the scoring procedure a graduate student at Harvard University specializing in psychology and with considerable experience in classifying thematic material, independently scored the responses to the EAT using the same criteria as the author.

Over-all agreement for all four categories is 48%. However, only two subjects were rated more than one category above or below that assigned by either rater. Pearson r correlating positions assigned by the two raters is .62.

The major discrepancies exist between ratings assigned to categories 1 or 2 and between 3 and 4. There are only 6 cases in which a subject assigned to either 1 or 2 is classified by the second rater as falling into 3 or 4.

Our next procedure was to establish two major categories of subjects: those who are "more prejudiced" and those who are "less prejudiced." We decided that all subjects assigned to categories 1 and 2 above would be classified as Group A (the less prejudiced group) and all subjects assigned to categories 3 and 4 above would be classified as Group B (the more prejudiced group).¹⁰ On this

¹⁰ Adjustments in ratings were made after the two raters consulted concerning their discrepancies. In all but one case the discrepant subjects were placed in Group A. This was done to ensure that the "more prejudiced" group would

basis, reliability of the two sets of ratings was again computed by chi square. With only 6 subjects classified in Group A by one rater and in Group B by the other, chi square is found to be 33.26 which is significant at the .001 level of probability.

Having established the reliability of the rating technique, we assigned our subjects to their respective groups. There are 42 children (70%), 27 girls and 15 boys in Group A; 18 children (30%), 10 boys and 8 girls making up Group B. The sex differences between the two groups are not significant and the sex distribution is also not significantly different from the total sex distribution for the group as a whole. Similarly there are no meaningful differences between the two groups on the basis of father's occupation, age, or religion. There are 12 Jewish children, 4 Catholic children, and 1 Protestant child in Group B.

Data on intelligence are available for 54 of our 60 subjects—38 in the less prejudiced group and 16 in the more prejudiced group. The subjects were administered the Pinter-Durost Intelligence Test, Scale I, Form A, as a routine part of the guidance program of the School Department, Town of Brookline.

The mean IQ for all 54 subjects was found to be 114.1 ± 11.2 . The range was from 85 to 140. Median IQ for the group was 112.5.

The mean IQ of our unprejudiced group is 115.5 ± 10.7 , while the mean IQ of the prejudiced group is 110.8 ± 12.4 . The critical ratio for the difference between these two means is 1.4 and is, therefore, not statistically different. The range of IQ scores in Group A is 96–140 and that in Group B is 85–134. The median IQ for Group A is 113.5 and the corresponding median for Group B is 110. These data are summarized in Table 6. The lack of difference in intelligence test scores between Groups A and B is similar to the finding of the California Study. In both investigations there is a slight tendency for increasing intelligence to be associated with a lack of prejudice, but this tendency does not reach statistical significance.

In the following sections, we examine in detail the thinking of our subjects with regard to problems that commonly form part of most intelligence tests. Since we have found no differences to exist between the general intelli-

have maximum homogeneity on the attitude dimension even at the expense of contamination of the "less prejudiced" group with persons of equivocal attitudes as judged by the divergence of the raters.

TABLE 6
RELATION OF IQ TO PREJUDICE RATING

Measure	Group A (N=38)	Group B (N=16)	Total group (N=54)
Mean IQ	115.5	110.8	114.2
Sigma	10.6	12.4	11.2
C.R.	1.4		
Median IQ	113.5	110.0	112.5
IQ Range	96-140	85-134	85-140

gence of prejudiced and unprejudiced children, we might expect that there should similarly be no differences between the two groups on tests that, for the most part, are likewise tests of intelligence. Should we find striking differences between the two groups on problems involving concept formation, deductive logic, and mental set not explainable from general intellectual factors, to what may we attribute performance failures on some tests of the higher mental functions? As a preliminary hypothesis, we suggest that, since prejudices involve the use of reasoning that by objective standards is faulty, those children in Group B who fail our tests or give poor performances in attempting to solve the conceptual problems given them are less intelligent than those in Group B who pass these tests. Implied in this hypothesis is the suggestion that even as early as the age of seven two types of "prejudiced children" are to be found: (a) those whose prejudice forms an integral part of personality, permeating all aspects of behavior and appearing on our tests as a failure to reason adequately, and (b) those whose prejudices are not deeply rooted in the personality but who have learned the culturally accepted modes of responses to various ethnic groups as part of their "home training" and express such learning through the stereotypes and clichés commonly found in prejudiced adults. We would suggest that children who are poor in abstract abilities, for example, who at the same time manifest ethnic prejudices are likely to be of the former "type" while children who are both adept at abstract conceptual thinking and show prejudices are likely to be among the latter "type." It is further suggested that it is that aspect of general intelligence rendering an individual capable of forming adequate generalizations or arriving at adequate conclusions that is generally depressed in children who are more prejudiced. While we have not attempted to item-analyze the intelligence test given our subjects, the fact that our tests of higher mental functioning are approximations of items used on standard intelligence tests will provide a

measure of the accuracy of these assumptions. Where supplementary independent evidence is available we have included it as a means of testing their validity.

Undoubtedly, some aspects of intelligence are lacking in children who engage in making broad and sweeping generalizations about individuals or groups. A qualitatively different variety of intellectual factors would appear to be required to obtain, for example, conditional disagreements with the items on ethnic groups on the EAT. The results of our tests reported below would seem to lend weight to the view that the significant intellectual component lacking in children who are high in ethnic prejudice is an inability to regard unknown or strange objects or events in terms of their intrinsic qualities or values. The ethnically prejudiced child appears unable to deal effectively with problems that are equivocal in nature. If this fact were generally true of the thinking of children high in prejudice, then their susceptibility to learn and to manifest ethnic prejudices may be a function of an incapacity to view any outgroup without apprehension. Similarly, the ethnically prejudiced child may be incapable of viewing an ingroup as possessing any but virtuous traits.

While "intolerance of ambiguity" may or may not be a general personality characteristic of the prejudiced person, it is most certainly a distinct cognitional aspect of prejudiced behavior. Evidence from the California Study and the extensive studies of Rokeach would appear to support the view that the central intellectual component of the prejudiced as opposed to the unprejudiced child is the relative inflexibility of the prejudiced child's thinking. Hence, although general intellectual ability shows negligible differences between high and low ethnic prejudice, the intellectual factor designated by the term "mental rigidity" appears to be associated with high prejudice and a relative "mental flexibility" seems associated with low prejudice.

We turn now to the problem of the relationship of prejudice to the functions of some of the higher mental processes. Our question will be: Are there significantly different relationships between the patterns of performance on tests of various mental functions that distinguish children high from children low in ethnic prejudice?

PREJUDICE AND INDUCTIVE REASONING

Concept formation or induction requires that specific objects or events be

so organized that a class of objects or events emerges. An individual may also hypothesize a concept, then test it on specimens from the class he assumes comprises the specimens. The former view of concept formation (or inductive reasoning as it is alternately known) is sometimes referred to as the "composite photograph" theory and the latter might be described as the "trial and check" theory, to adapt a phrase from Woodworth (60).

Concept formation in children has been regarded as a mental function that makes its appearance approximately between the ages of seven and eight.

Piaget's work with Swiss children (39) found in general that there are stages of development in reasoning. He maintained that reasoning below the age of 7 or 8 remains on the level of "syncretistic ego-centrism" and the child has great difficulty in making objective generalizations. Since his work, considerable evidence contradicting his position has appeared (6, 16, 18, 24, 28, 54).

Many theorists now hold that reasoning is a gradually emerging function that does not change by stages but rather by degrees. The seven-year-old child may be capable of handling simple concepts in a fashion similar to that employed by older children or adults. Moreover, a variety of reasoning "types" may be found at every age level. "Typical" reasoning presumed to characterize the thinking of young children concerning natural phenomena were found also at the adult level (37). Isaacs (24, p. 92) in her critique of Piaget sums up her work by pointing out that she found no evidence of mental "structures" that change in nature with age, but that

"... They show, rather, a continuous advance in scope and clarity of neotic synthesis and in the ability to handle experience in more and more complex form. The spontaneous behavior of these little children reveals the same continuity of growth as Burt's graded tests of reasoning have demonstrated in children from seven years of age onwards."

We tend to concur with the views of Isaacs and consequently have assumed that reasoning "ability" in our subjects will vary in degree rather than in kind.

With the foregoing assumption in

mind, we may define inductive reasoning as that process of thought in which the individual attempts to arrive at a general concept relating particular events or objects as members of a given class of events or objects.

To test this process of thought we used a group of eight Weigl-type objects (57) in a situation similar to that presented by the Wisconsin Card Sorting Test for Adults. These objects consisted of two groups: a "large" group of four objects measuring 5.5 square inches each, containing a red square, green circle, white triangle, and red triangle; a "small" group contained four objects of about 3 square inches. These were a blue triangle, a red square, a white circle, and a blue square.

Large Group

Red square ($2 \frac{11}{16}'' \times 2 \frac{11}{16}''$)
Green Circle ($1 \frac{15}{16}''$ diameter)
White triangle ($3 \frac{1}{2}''$ equilateral)
Red triangle ($3 \frac{1}{2}''$ equilateral)

Small Group

Blue triangle ($2 \frac{1}{8}''$ equilateral)
Red square ($1 \frac{1}{2}'' \times 1 \frac{1}{2}''$)
White Circle ($\frac{3}{4}''$ diameter)
Blue square ($1 \frac{1}{2}'' \times 1 \frac{1}{2}''$)

Consequently, the objects comprise four colors, three shapes, and two sizes. The objects were made of ordinary desk blotter cut to size.

Procedure

The subject was seated and the instructions given by *E* while he removed the objects from an envelope and scattered them on the table before *S*.

"Today we have a puzzle to do. Do you like puzzles? Good. Now, do you see all these different things on the table? Well, let's count them first and see how many there are. (*E* allows *S* to count the objects and assists him if necessary until *S* knows that there are eight objects.) Now here is what you have to do to get this puzzle. You must put these different things into groups, just like the groups in your class, so that the things that belong together go together in the same group. In each group must be the same number of things and only things that belong together must be in each group."

If the instructions were not understood, the *E* repeated them slowly and in a few cases *S* was urged to try to put the objects together (*E* employing some of the key words used in the instructions) as he began to deal with them.

The solution of the task set for *S* is to group the objects into one of two possible arrange-

ments: (a) Two groups, one of "large" and one of "small" objects; or (b) Four groups consisting of two "large" and two "small" subgroups. The subject, to be credited with a solution, was required to verbalize his solution—to make it plain that the size of the objects was the critical concept.

A maximum of 15 minutes was allowed for the procedure. Four hints were given the subject, if required, at intervals during the testing session. The exact point at which a hint was introduced varied from subject to subject depending upon the state of the *S*'s progress toward a solution. Typically, when *S* ceased concentrating on the problem or seemed at an impasse, a hint was offered.

If a hint was given and produced no renewed effort in *S* to go on with the task, the next hint was offered. The hints were as follows:

Hint 1: The large red square and large white triangle were placed apart from the rest of the objects and *S* was told, "To get this puzzle, this one and this one (pointing) must be in the same group. They belong together."

Hint 2: The first hint was placed before *S*. Then the small blue triangle and the small red square were placed apart from the rest of the objects and *S* was told, "To get this puzzle, this one and this one (pointing) must be together in the same group also. They belong together, too."

Hint 3: The two previous hints were placed before *S* and *E* picked up the small blue square and placed it in the "small" group, saying, "This one belongs in this group."

Hint 4: The *E* placed the three previous hints conspicuously before *S* and reminded him that these objects belonged together in the two groups as indicated above. The *E* then took the large green circle and placed it with the "large" group, saying, "This one belongs in this group; now what shall we do with the rest?"

With Hint 4, *S* had only two remaining objects: the small white circle and the large red triangle. If, after the fourth hint was given, *S* continued to remain in a quandary, *E* "tested the limits." The solution was presented to the *S* in a 4×4 grouping, and he was asked to tell why this grouping solved the puzzle. If this failed, *E* then made an individual comparison of one small with one large object, asking "How are these two the same?" and "How are they different?" If *S* failed this, the test was discontinued. If *S* made the size comparison after this final procedure, the objects were then regrouped by *E* into a 4×4 size solution and *S* was asked, "How are all these the same?" (pointing first to the small group, then to the large group). Finally *E* passed his hand over all of the objects in the large group and said, "How are all these different from all these?" (pointing then to the small group).

One further hint was employed. If *S* was still unaware of the nature of the solution after three hints were given, an "arithmetic hint" was given. The *E* asked the subject to tell him how many and how many made 4. If *S* replied, "Two and two," *E* insisted again until a "three and one" response was forthcoming. Then *E* pointed to the Table and asked *S* to count the number of objects in each "hint" group and in the remaining group.

Results

Of the 58 subjects who took the concept formation test, 24 or 41.3% achieved "spontaneous" solutions. Such solutions on the average required about eight minutes of work and about two hints. Three subjects required no hints at all, 9 subjects required two hints, 1 required three hints, and the remaining 11 subjects required four hints.

Eighteen subjects failed to achieve any form of solution to the problem. This number represents 31.1% of the total group. All of these subjects were tested through the "testing-the-limits" procedure described above, each received the four standard hints and each was presented with the correct solution but could not comprehend its nature.

Sixteen subjects, or 27.6% of the group, achieved solutions that were either inaccurately verbalized or were based upon a concept other than size or were a consequence of the "guiding" procedure employed past the four regular hints. Hence we would prefer to call the solutions achieved by this latter group as "guided solutions."

Although only 40 subjects achieved some form of solution, 45 subjects or 77.6% of the group made at least some use of the concept of size at some point in their manipulation of the objects. A variety of solution and failure patterns of concept formation was found.

The use of shape seems to predominate in both the cases of subjects who solved and those

who failed to solve the problem. The pattern achieving the greatest number of solutions (shape, color, and size) and that achieving the second greatest number of solutions (shape and size) contains no reference to the use of "ideas." By *ideation*, we refer to the child's use of the geometric objects as elements in a design or "picture." For example, he would place several objects together and announce: "A tree with a trunk" or "a face with a hat." Ideational responses involved from two to all eight objects. Burt (9) and others have described the creation of designs in situations demanding attention to other, more abstract elements, as a more primitive type of response than attention to color or shape. In order of increasing abstractness, the determinants are: ideation, color, shape, and size. We shall consider each of these determinants in turn before analyzing the results in greater detail.

What we have called "ideation" would appear to be the natural play-tendency that must be temporarily suppressed while the child attends to the rules of the new "game." Design-building is the least abstract response to the situation simply because it is the most concrete in a situation demanding abstraction. Hanfmann and Kananin (19) have similarly described a "personification" response tendency in schizophrenics as a primitivization of the thought process.

Color and shape are next in order of abstractness. Revesz (42) and Thompson (54) both have found color responses to be more typical of the very young children, while shape responses appear later in genetic sequence. In dealing with the objects both shape and color are potential "concepts." The individual may organize the objects into a variety of such concepts. The concepts thus derived are primarily a function of their stimulus character: shape and color are the most obvious surface characteristics. Shape and color concepts may here be referred to as "first order" concepts. They stem directly from the nature of the stimulus objects. The concept of size (the solution concept of this study is a "second order" concept: the concept derives from the *relationships between the objects*. To achieve this second-order concept, the individual must (a) avoid design-building, (b) ignore the surface character of the object (shape and color) as irrelevant to the solution, and (c) note the differences between two sets of objects.

We now turn to the central problem of this chapter: Do the methods employed by more prejudiced children in reasoning inductively markedly differ from those employed by less prejudiced chil-

TABLE 7
ANALYSIS OF RESULTS OF CONCEPT FORMATION
TEST ACHIEVED BY GROUPS A AND B

	Group A		Group B		Total	
	N	%	N	%	N	%
Spontaneous solution	19	46.3	5	29.4	24	41.4
Guided solution	14	34.1	2	11.8	16	27.6
No solution	8	19.6	10	58.8	18	31.0
Total	41	100	17	100	58	100

dren? We have hypothesized that the methods differ in their "adequacy" in achieving a solution to the problem. One method of testing the adequacy of the methods employed by both groups is to determine whether or not solutions were, in fact, actually achieved. Obviously, arrival at the concept of size implies the employment of "adequate" methods.¹¹

In Table 7 three categories of "solution" are presented: Spontaneous solution, guided solution, and no solution. A spontaneous solution is a solution achieved through the individual's efforts aided only by the standard hints. A guided solution is a solution achieved by hints beyond the four standard hints in the course of "testing-the-limits." We do not mean to imply that the individual does not achieve a genuine understanding of the size principle as the "solution" but only that S received additional help and was guided to look for a concept other than those he had spontaneously employed. No solution resulted when the individual could not arrive at a solution despite all attempts at "guidance" by E in the "testing-the-limits" situation.

The ratio of spontaneous solutions comparing Groups A and B is 19/5 and is statistically significant. It should be noted that only 24 cases are involved. If

¹¹ Actually, this is not necessarily the case. Some subjects arrive at the solution in ways defying description. The solution simply appears, but the process of achieving it is not apparent to E, and it is a mystery to S, who is surprised to learn that the game is at an end.

we include the "guided" solutions with the spontaneous solutions, we find that the ratio of solutions to failures comparing groups A and B is, of course, also significant. Thus, it would appear that *more prejudiced children are less capable of arriving at the appropriate concept spontaneously and/or with additional aids from E*. As it might be expected, the number of no-solutions in the A group is significantly smaller than the number of no-solutions in Group B. Hence, there appears to be a marked inability among the more prejudiced children to arrive at the appropriate conclusion than among the less prejudiced children. Moreover, the ratio of 19 spontaneous solutions to 8 failures (for Group A) is significantly better (in terms of the task) than the ratio of 5 spontaneous solutions to 10 failures (for Group B).

We have partial proof that the ability to reason inductively is associated with a lack of ethnic prejudice and that inadequate inductive reasoning is associated with the presence of ethnic prejudice.

It may be argued that these findings do not demonstrate the relationship of conceptual difficulty among prejudiced children but do, in fact, indicate that the children in Group B are simply less intelligent than those in Group A. Similarly, a larger proportion of intelligent children may be found in Group A than in Group B. We may dismiss this argument at once, since we have shown that there is no significant difference in intelligence (as measured by IQ) between Groups A and B.

A more challenging case might be argued that those who passed the test (i.e., achieved a size solution) were more intelligent than those who failed. However, biserial r correlating intelligence and passing or failing is found to be .07,

indicating that intelligence, as such, is not the crucial factor in accounting for the differences between the two groups.

One approach to the qualitative analysis of our data is to determine "how close" the subjects who failed were to a final correct solution. Although 40 subjects passed this test, 45 subjects had made use of the concept of size at some point in their dealings with the objects. Closer inspection reveals that 4 of the 5 additional subjects using size are among the 8 in Group A who failed and 1 is from the no-solution subgroup of Group B. We may infer the adequacy or inadequacy of an approach to the objects in terms of their apparent relevance to the solution. We shall employ the term "hypothesis" to refer to the suggested uses offered by the subject in dealing with the objects. Some hypotheses are inferred from the subject's performance when no verbalization was forthcoming.

Comparing first the total number of "hypotheses" offered by the children in each group, disregarding whether or not a solution was reached, we find that there are 197 hypotheses in Group A with a mean of 5.05 ± 1.69 . There are 99 hypotheses offered by Group B; the mean numbers of hypotheses is 4.95 ± 1.44 . The difference between the two means is not significant. Do those who achieve solutions offer a greater number of hypotheses than those who fail? What is the relation of hypotheses in solution and no-solution groups in both the A and B groups? To answer these questions, we calculated the differences between the mean number of responses for each subgroup. Table 8 summarizes the findings.

We find that there are no differences between the number of hypotheses offered by less prejudiced children regardless of the outcome of their efforts. How-

TABLE 8

TESTS FOR THE DIFFERENCES BETWEEN MEANS OF SOLUTION AND NO-SOLUTION SUBJECTS IN GROUPS A AND B

	Group A			Group B		
	Solution	No Solution	Total	Solution	No Solution	Total
<i>N</i> Hypotheses	154	43	197	31	68	99
Percentage	78.2	21.8	100	31.3	68.7	100
Mean Hypotheses	4.7	5.4		3.9	8.5	
Sigma	1.97	1.40		1.17	1.89	
<i>T</i>	1.13			5.75*		

* Significant beyond .01 level of confidence.

ever, a large and significant difference appears between the number of hypotheses of the more prejudiced group: no-solution subjects offer a great many more hypotheses than do subjects who reach a solution. What of the relationship of Group A solution to Group B solution and Group A no-solution to Group B no-solution?

We find a significantly greater number of hypotheses given by Group A-solution than by Group B-solution. But there are a significantly greater number of Group B-no-solution hypotheses than Group A-no-solution hypotheses. The above results suggest that (a) less prejudiced children take greater pains in achieving a solution by promulgating a greater number of hypotheses toward the correct solution when they achieve it; (b) More prejudiced children tend to attempt fewer guesses concerning the correct solution before they achieve it than do less prejudiced children; but (c) more prejudiced children who fail to arrive at a conclusion tend to offer a great many more hypotheses than either the less prejudiced solution or no-solution groups or the more prejudiced solution group. If the gross number of hypotheses is taken as an indication of the pertinence of the approach, it would appear that more prejudiced no-solution children develop

more adequate means for arriving at a solution despite the fact that they fail to do so. Why then do they fail? Below we offer an analysis of the number and type of inadequate hypotheses employed by each subgroup. Our criteria for inadequacy were:

1. "Personifies" the objects or makes a design of them.
2. An arbitrary concept is offered i.e., "The reds are really blue."
3. Repeats an hypothesis found lacking previously.
4. Attempts to make a jigsaw puzzle out of the objects.
5. Suggests altering the shape or color of objects.
6. Blindly matches the *E*'s hints with objects.
7. Assumes that there are too few or too many objects.
8. Random shifting of the objects occurs.
9. *S* accuses *E* of hiding objects or of trying to "trick" him.

From an objective point of view, these types of approach would seem unlikely to lead the *S* to a solution but rather would tend to lead him astray from the main task. It will be noted that not all of the items are "hypotheses," strictly speaking. We have used this term in a generic sense as also meaning "trial" or "attempt." In Table 9 we present the results of the analysis of inadequate hypotheses for each of the four subgroups.

At first we note that more than half

TABLE 9
TYPES OF INADEQUATE HYPOTHESES GIVEN ON CONCEPT FORMATION TEST

Hypotheses	Group A				Group B				Total	
	Solution		No Solution		Solution		No Solution			
	N	%	N	%	N	%	N	%	N	%
Designs	15	25.0	3	10.7	3	17.6	12	22.6	33	20.9
Arbitrary concepts	4	6.7	4	14.3	3	17.6	6	11.3	17	10.7
Color repetition	3	5.0	1	3.6	2	11.8	3	5.7	9	5.7
Shape repetition	26	43.3	14	50.0	2	11.8	20	37.7	62	39.2
Jig-saw fitting	4	6.7	1	3.6	1	5.9	3	5.7	9	5.7
Change, shape or color	3	5.0	2	7.1	2	11.8	1	1.9	8	5.1
Hint-matching	2	3.3	1	3.6	2	11.8	1	1.9	6	3.8
Assumes fewer or more objects	3	5.0	1	3.6	2	11.8	4	7.5	10	6.3
Random shifting	0	0.0	0	0.0	0	0.0	2	3.8	2	1.3
Accuses <i>E</i> of hiding objects	0	0.0	1	3.6	0	0.0	1	1.9	2	1.3
Total inadequate	60	100	28	100	17	100	53		158	100
Total responses	154		43		31		68		296	
Percentage inadequate		39.0		65.1		54.8		77.9		53.4
Total inadequate		88				70				
Percentage of total <i>N</i> responses		44.6				70.7				

of the 296 hypotheses offered by the total group of subjects are "inadequate" according to our definition. In the A group, there are 88 inadequate hypotheses or 44.6% of the total responses in Group A. Group B produced 70 inadequate responses of 99 offered, or 70.7% inadequate hypotheses. This difference is reliable. Hence, *more prejudiced children tend to produce more inadequate (or fewer adequate) hypotheses in attempting to arrive at a second-order concept than do less prejudiced children.*

We further find that the difference of 15.9% fewer inadequate hypotheses separating Group A-solution and Group B-solution is not significant although it is in the indicated direction. It is felt that the unreliability of this difference stems from the fact that there are 123 fewer responses among the B-solutions than among the A-solutions. This latter discrepancy may be traced to the small number of subjects in Group B who solved the problem (eight subjects) and the comparatively few hypotheses offered in their solutions (31 hypotheses). There

are other evidences of more adequate responses among the less prejudiced children: (a) the frequency of inadequate hypotheses in the B-no-solution group is significantly larger than that of the A-no-solution group. Hence, even in failing to solve the problem, less prejudiced children make "better" attempts toward its solution.¹² (b) In the two types of inadequate hypotheses occurring most frequently (designs and shape repetition), the A group suffers significantly fewer failures than does the B group, despite their occurrence. Apparently, designs and shape-repetition cannot be converted to advantage by the B group as it may be by the A group.

Bringing together the findings, we note that, although the mean number of hypotheses for Groups A and B do not differ:

1. The mean number of hypotheses of Group A-solution differs importantly

¹² As might be expected, the no-solution groups in each of Groups A and B make significantly more inadequate attempts to solve the problem than do their corresponding solution groups.

from the mean number of Group B-solution hypotheses. While the pertinence toward solution of the hypotheses offered by each group does not differ significantly, there are indications of qualitative differences in the types of hypotheses offered (as well as in the frequency of inadequate hypotheses in the B-solution group) in the direction of the solution in the more prejudiced group.

2. The greater frequency of inadequate hypotheses and the relatively small number of total hypotheses offered by the B-solution group suggests that more prejudiced children tend to "jump to conclusions" with fewer correct leads, even though a solution ensues. Less prejudiced children appear to take more guesses about the nature of the solution and these guesses seem to be more reality-bounded as well as less random.

3. Even when failing to arrive at the "solution" to the problem, less prejudiced children attempt fewer hypotheses, which are, however, more adequate to the achievement of a solution, than do more prejudiced children.

4. Where perseveration occurs, less prejudiced children appear to turn it to advantage (perhaps this fact would be more correctly interpreted as indicative of "stewing" or "mulling" among the less prejudiced children who finally come out with a solution). The more prejudiced children who appear to perseverate fail to achieve a solution in the ratio of about 10/1.

On the basis of the above analysis we are now prepared to assert that hypothesis 1 has been substantiated by our findings: in situations requiring the formation of general concepts . . . more prejudiced children tend to be deficient in forming adequate concepts, less prejudiced children tend to develop adequate concepts.

PREJUDICE AND DEDUCTIVE REASONING

Most studies of reasoning do not, as a rule, specify what area of reasoning or of thought is to come under investigation. Rather, the reader is left to discover the types of reasoning studied from the context. Often, no attempt is made to distinguish between induction and deduction. As Woodworth has pointed out (60), induction and deduction are not necessarily different processes, but they do involve the manipulation of *different types* of material. Whether or not the processes involved in the solution of inductive and deductive problems are different would seem to be an empirical problem. No systematic attempts, however, have yet been made to establish the identity or uniqueness of these processes. There does seem to be an implicit assumption of difference in the two types of problem-solution in many intelligence tests for children.

Strictly deductive problems typically begin with the enunciation of a premise or set of premises. The individual is to "know" certain facts or to "grasp" certain relationships before the introduction of a question concerning "the next logical step" in consequence of the preceding steps. Sometimes, due to a desire to standardize the procedure and to eliminate extraneous associative factors, symbols are used in place of words in measuring or analyzing the reasoning of an individual. Hence, formal syllogistic logic has tended to become the medium through which this measurement or analysis has been carried out.¹³ The studies of Wilkins (59), Woodworth and Sells (61), Sells (50), Sells and Koob (51), and O'Connor (38), with adults as subjects, has taken the form of the solution of syllogistic problems. With children, Burt (9) employed syllogistic material in his "graded tests of reasoning" with great success. He considered these tests as the best single indicators of intelligence among older children. Although the items he employed are not strictly syllogistic,

¹³ It is not implied that deduction has been studied primarily through syllogistic logic. Rather we wish to point to the growing tendency to carry out empirical studies using syllogisms rather than other materials.

they have the general characteristics of the syllogism. Similar items appear in a variety of intelligence tests.

In the present study, we employed a series of items similar to those of Burt, but with certain changes introduced in the *form* of the items. The Deductive Reasoning Test (referred to hereafter as the DRT) is composed of nine items, an introductory item to familiarize *S* with the nature of the task, and eight test items. The final nine items selected were:

1. Electric lights never twinkle. This light is twinkling. Is it an electric light? Why?
2. It runs but it has no feet; it roars but it has no voice. What is it? Why?
3. Jane is taller than Jill. Mary is shorter than Jill. Who is the tallest of all?
4. You're going to the store. On the right-hand side is a farm, straight ahead is a lake. Which way is it to the store? Why?
5. The green bug runs faster than the red bug, the red bug runs faster than the blue bug. Which bug is slowest of all?
6. When it's hard you can throw it; when it's soft you can fall into it; you can swallow it, you can freeze it, and you can boil it. What is it? Why?
7. There are three babies. Johnny is older than Janie; David is younger than Janie. Which baby is the oldest of all?
8. Little boy A is short and fat. Little boy B is tall and thin. Little boy C is tall and fat. Which boy weighs the most?
9. If you touch it in the summertime it's cold. If you touch it in the wintertime, you may be burned. It's made of iron and its sometimes painted silver. What is it? Why?

After *S* was seated he was instructed as follows:

"Today we're going to have riddles. Do you like riddles? Good. Now listen carefully to each riddle because I can only say them once. See if you can guess each riddle."

The problems were then presented in turn. The test took about 20 minutes per child to administer. Two types of problems were presented to the child: (a) problems that could be answered with only one correct response. These are problems of straightforward "formal logic" (items 3, 5, 7, and 8); (b) problems that could be answered correctly with any one of a number of possible responses ("ambiguous" problems including items 2, 4, 6, and 9). The items cover a wide range of difficulty and are of varied complexity. The ambiguous items were included to test subsidiary hypotheses suggested by the work of Frenkel-Brunswik (14) that prejudice is correlated with (a) an inability to cope with ambiguous stimuli in a flexible manner, and (b) the inability to take into account the multiplicity of potential responses. Rather, in a series of

TABLE 10
CORRECT, PREFERRED, AND INCORRECT
RESPONSES TO SEVEN ITEMS OF THE
DEDUCTIVE REASONING TEST

Items	Correct and Preferred		Incorrect		Total	
	N	%	N	%	N	%
2	31	58.5	22	41.5	53	100
3	49	86.0	8	14.0	57	100
5	48	82.7	10	17.3	58	100
6	67	81.7	15	18.3	82	100
7	36	64.3	20	35.7	56	100
8	23	40.4	34	59.6	57	100
9	19	33.3	38	66.7	57	100
Totals	273	65.0	147	35.0	420	100

studies on memory and perception, Frenkel-Brunswik and her co-workers found that more prejudiced persons tended to select one of the potential responses and to cling to it tenaciously in spite of the changing stimulus field.

As used in this study, deductive reasoning is defined as that process of thought in which implications are drawn from given premises. Based upon this definition, the results have been analyzed in a number of different ways. A total of 473 positive responses were obtained in answer to the 8 test problems, an average of 58.6 responses to each problem. Responses were obtained from 57 of our 60 subjects (40 in the less prejudiced group and 17 from the more prejudiced children). Table 10 presents a summary of correct and incorrect responses to both formal logic and ambiguous questions. One may question whether there can be a "correct" response to an ambiguous question. However, the ambiguous items are so worded that a "preferred" response might be given. In the case of item 4, none of the three obvious alternatives need necessarily be "correct." All are potentially acceptable if the reasons offered as support for choosing one of them are adequate. Hence, the "correct" responses to the ambiguous items are in reality only *preferred* (most pertinent to the premises) responses. The results for item 4 are presented separately below in a more extended analysis.

Two-thirds of all the responses, regardless of type of item, are correct or preferred. Some items are obviously more difficult than others. The "easy" items would appear to be items 3, 5, 6, and 7. Only a third of the responses to item 9 are correct; item 8 was correctly re-

TABLE 11
RESPONSES TO DRT ANALYZED INTO
FORMAL-LOGIC AND AMBIGUOUS
CATEGORIES

	Ambiguous		Formal logic		Totals	
	N	%	N	%	N	%
Correct (preferred)	117	60.9	156	68.4	273	65.0
Incorrect (nonpreferred)	75	39.1	72	31.6	147	35.0
Totals	192	100	228	100	420	100

sponded to in 40% of the cases.¹⁴ Item 2 is correctly responded to in 63% of the cases.

Analyzed into ambiguous and formal logic categories the questions are responded to in the following fashion:

Chi square¹⁵ for the differences in type of responses to the two types of problems is not significant at 2.71. The ratio of ambiguous to formal logic correct responses indicates that in general the more ambiguous problems are more difficult than are the formal logic problems.

To return for a moment to item 4, we find that of the 53 responses to this problem, 25 or 47.2% are responded to in terms of the "atmosphere" of the problem (the store is said to be to the left). The remaining 28 responses or 52.8% are distributed between responses of "to the right" or the "farm-way" and "straight ahead" or the "lake-way." None of these responses may be considered incorrect, although the rationale for choos-

¹⁴ It should be noted that item 8 is more ambiguous than was intended. A good many children saw no reason why a short, fat boy should weigh less than a tall fat boy. Nevertheless, a substantial number asserted the relation of height and girth to weight in the intended fashion.

¹⁵ This analysis is based upon one response per subject for the most part, hence chi square is employed.

ing one of them varies greatly in acceptable reasoning from subject to subject. The reply "to the left" may be based on the subject's understanding that the farm and the lake stand as *barriers* or as illogical places to have a store. Some subjects derived the response from "pure logic": left was the only *forward* direction not mentioned in the problem. Many subjects felt as did one that: "it could be anywhere." Here, the reasoning was that the premises did not necessarily limit the location of the store to any particular place in any particular direction. One subject replied: "It might be behind you or it might be in town, wherever the town is." An incorrect (nonpreferred) response is the following: "It's straight ahead." (Why?) "The store is in the lake 'cause the lake has no water in it." Considering the ambiguity of the problem, the store might conceivably be in the lake if it were dry, although reality demands that we label this reasoning as "inadequate." We shall come back to the problem of item 4 at a later point in the analysis.

We now are prepared to test the hypothesis that "more prejudiced children tend to produce more numerous as well as more invalid conclusions than do less prejudiced children."

A cursory examination of the results summarized in Table 12 indicates that at least the first part of the hypothesis (that dealing with the number of conclusions) is not verified. Although all subjects did not give usable responses to each problem, we find that the mean number of responses given by the Group A for seven items is slightly more than 1.0 per subject and for Group B the mean is 1.13 per subject. The difference is insignificant. The fact that S gave only one response to each question (item 6 is a notable exception) may be in part due

TABLE 12

NUMBER AND TYPE OF RESPONSE TO DEDUCTIVE REASONING TEST BY MORE AND LESS PREJUDICED CHILDREN

Item	Type of Response ^a				Totals
	Group A		Group B		
	Cor- rect	Incor- rect	Cor- rect	Incor- rect	
2 ^b	23	8	8	14	53
3	36	4	13	4	57
5	32	9	16	1	58
6 ^b	47	5	20	10	82
7	24	15	12	5	56
8	14	26	9	8	57
9 ^b	16	23	3	15	57
4	16	24	9	4	53
Totals not including Item 4	192	90	81	57	420

^a Don't Know responses are not included in this table, since we wish to consider here only positive responses to the problems. We shall consider the meaning of Don't Know responses at a later point.

^b Ambiguous problems.

to the nature of the problems. In those involving linear relationships of increasing size or age, little speculation is possible. The ambiguous problems offer a wider range of possible answers. Hence, speculation would appear more frequently. However, the more prejudiced children are not more speculative than the less prejudiced children in response to either the formal logic or the ambiguous problems.

The latter part of the hypothesis we are to test deals with the validity of the conclusions reached by S. We first compared the correct (preferred) and incorrect (nonpreferred) responses of all subjects to seven items of the DRT (again omitting item 4) and found the differences between Group A and B to be not significant. Hence, correctness or incorrectness of the responses as a whole shows no tendency to differentiate prejudiced from unprejudiced children.

TABLE 13

PREFERRED AND NONPREFERRED RESPONSES ON AMBIGUOUS PROBLEMS BY MORE AND LESS PREJUDICED CHILDREN

Type of Response	Group A		Group B		Total	
	N	%	N	%	N	%
Preferred	86	70.5	31	44.3	117	60.9
Nonpreferred	36	29.5	39	55.7	75	39.1
Total	122	100	70	100	192	100

We next compared the results of the two prejudice groups against the two types of problems. Chi square for the formal logic problems is 1.10 and is insignificant. Chi square for the ambiguous problems is, however, 11.39 and is significant beyond .001 level of confidence. (Table 13) Even if item 4 is included in the computation of the results and the "left side" response is taken as correct, p is still at the .02 level of probability. It appears that the accuracy of deduction is dependent upon the nature of the problem to be solved: the more prejudiced children having much more difficulty in arriving at valid conclusions on problems involving some ambiguity than do less prejudiced children. Both groups, on the other hand, have equal success (or difficulty) on problems of formal logic.

Group A, on both types of problems, succeeds on two problems for every one that it fails. In Group B, on the other hand, the ratio of correct to incorrect on formal logic problems is about 3/1, but on the ambiguous problems the ratio is about 1/1 leaning in the direction of more nonpreferred responses.¹⁶ These dramatic differences would seem to confirm Frenkel-Brunswik's "intolerance of ambiguity" hypothesis: The prejudiced

¹⁶ Statistically, chi square for the differences in success in Group A for each type of problem is .62 and p is $< .50$. The comparable chi square for Group B is 31.65 and p is beyond the .001 level.

children in our study are far less capable of dealing in logical fashion with ambiguous problems than they are capable of succeeding with clear-cut formalistic problems. These findings are consistent with those of O'Connor (38), who studied the relationship of abstract reasoning ability to ethnocentrism, ethnic prejudice, and intolerance of ambiguity in 77 Harvard undergraduates. She found ethnocentrism related to difficulty in abstract reasoning when intelligence was controlled. Further, intolerance of ambiguity was related to poor abstract ability *only* when associated with ethnocentrism. In view of the argument to follow, we would interpret O'Connor's findings to mean that all three factors are related to some unknown prime causative factors. These factors produce the prejudiced approach to life in both the world of objects as well as social relations.

An important factor in the differences obtained are the comparatively large number of nonpreferred responses of Group B. In order to determine the nature of these responses and their comparative likenesses and differences between more and less prejudiced children, we undertook an item analysis of the responses to ambiguous problems. Before presenting the results of that analysis, we should like to return to the formal logic problem for a moment.

One may ask, "What is the nature of the task for *S* in the so-called formal logic problems?" The *S* is required to compare two sets of functions (*a* is older than *b*; *b* is younger than *c*) and then to derive the logical relationships holding between both sets of ages of *a*, *b*, and *c*. Only one sequence is possible: there is only one correct deduction to be made. Moreover, the correct response is present as part of the statement of the problem. This sort

of problem is not difficult for young children. Nearly 75% of Burt's subjects (9) who passed such items (his *Ss* ranged from 5 to 12 years) and 80% of *Ss* in the present investigation found little trouble in selecting the correct response.

The ambiguous items pose a different kind of problem to the child. There is no simple choice to be made. The *S* must select the one reply that most closely fits the premises. His response must reside in his ability to understand. Moreover, as in items 2, 5, and 10, the "something" that the *S* must deduce has certain qualities which seem to be anachronistic or self-contradictory. In item 10, for instance, the child searches for an object whose thermal characteristics are inversely related to the outside temperature. In item 6, he is asked to deduce something that seems capable of changing like the proverbial chameleon. In item 2 he must find something that behaves in animate fashion but does not possess animate attributes to account for this behavior. Hence, these problems are more complicated and more baffling to young children, to judge by our results. But of special interest is the fact that less prejudiced children have reliably less difficulty in dealing with these ambiguous problems than do more prejudiced children. To analyze this finding in detail we examined the nature of the responses to two of the ambiguous problems.

We have raised the problem presented by item 4 and discussed some of the general results obtained. Three types of responses are made to this problem.

A further analysis of each type of response reveals that those *Ss* who choose "left" as their conclusion in almost every case create what amounts to a formal logic problem of this item. The reasoning is somewhat as follows: "To the right

TABLE 14
RESPONSES TO ITEM OF DRT BY
GROUPS A AND B

Response	Group A		Group B		Total	
	N	%	N	%	N	%
Left	16	40.0	10	58.8	26	45.6
Farm-way	16	40.0	4	23.5	20	35.1
Lake-way	7	17.5	1	5.9	8	14.0
Other	1	2.5	2	11.8	3	5.3
Total	40	100	17	100	57	100

is a farm, so there is no store *there*. Straight ahead is a lake so there is no store *there*. Therefore, there is only one place for the store to be and that is to the left." Hence, the lake-way and the farm-way are "blocked" or "occupied." The "unblocked" or "unoccupied" side is chosen as the only alternative. For these children, this is not a choice but an elimination problem, having the elements of formal logic problems. The solution of the "riddle" resolves into the following syllogistic format: given A, B and C; it is not A and not B, it must, therefore, be C. Some examples of this reasoning follow:

"Left. 'Cause you didn't say left."

"The left is the store, 'cause it can't be in the farm or up by the lake."

"Left. There is nothing else." (Must it be on the left?) "Yes, the lake is in front and the farm is right."

"Left. 'Cause farm's on the right, lake ahead. Store left."

"The left-hand side. You left out left, the others have something."

Those who choose the lake-way or the farm-way are not restricted by the "logic of the situation." Rather they view the problem as an indefinite one. For some, the farm is not an impediment and for others the lake is not a barrier. The locations of the lake and the farm do not necessarily determine the location of the store. Some typical examples are:

"Right hand. Oh, I meant straight ahead." (Why?) "'Cause it could be near the lake." (Must it be near the lake?) "No, could be near the farm if the farm and the lake are near one another."

"Towards the lake. You can't buy things at the lake—you can buy things near the lake."

"Near the farm. 'Cause farms are near stores and lakes are further."

"To the barn (farm). In a lake you can't buy food." (Is store in the farm?) "Could be or could be on the other side of the farm."

"Might be lake-way. Might be in the park or somewhere. You can walk past the pond and the store would be a little bit further on."

Since the location of the store is ambiguous to these children, room is left for imagining where a store might or could be. The child calls upon experience and imagination to fill the gap left by the ambiguity of the problem. It is important to note that the formal logic solutions do not permit alternatives. It *must* be "to the left" because "nothing is left" to choose. Thus when asked, "Could it be somewhere else?" *none* of the formal logic children reply affirmatively. On the other hand, 14 or 50% of Ss giving other-than-left responses admit the possibility of the store's being in a number of different places. Since all of the children were asked to state whether the store *could* be located elsewhere, the above figure indicates at least the minimum number admitting the possibility of an alternate locale.

Examining the responses of Group A and B according to the types of reasoning involved in reaching a conclusion (eliminating uncertain and erroneous responses) we find that there are a significantly larger number of non-formal-logic reasons given for Group A than for Group B responses. Table 15 summarizes these findings.

Hence, less prejudiced children tend to allow a degree of flexibility in their interpretation and response to item 4 while more prejudiced children tend to

TABLE 15
REASONS OFFERED TO EXPLAIN CONCLUSION
DRAWN IN ITEM 4 OF DRT BY
GROUPS A AND B

Reasons	Group A		Group B		Total	
	N	%	N	%	N	%
Formal	14	37.8	10	71.4	24	47.1
Nonformal	23	62.2	4	28.6	27	52.9
Total	37	100	14	100	51	100

interpret the problem in a somewhat rigid manner and respond to it in formalistic terms.

Item 6 of the DRT concerns a subject having six qualities. The S is required to keep in mind its different properties. Six response categories were employed in analyzing the data: water, food, ice or snow, mud or sand, an object, don't know.

What is a *valid* conclusion to be drawn from this problem? The subject of the problem has plasticity, since it can exist in both a hard and soft state, it can be ingested and is capable of being subjected to extreme temperature changes. Perhaps its key property is its capacity for being swallowed. This suggests a food of some sort. We do not generally fall into food, however, although this is entirely possible in the reasoning of small children, to judge by the results. We can both fall into snow and swallow it but we generally do not boil it. Snow is made

TABLE 16
ANALYSIS OF CONCEPTUAL AND FURCATED
TYPES OF RESPONSE TO ITEM 6 OF DRT

Response	Group A		Group B		Total	
	N	%	N	%	N	%
Conceptual	31	83.8	7	46.7	38	73.1
Furcated	2	5.4	6	40.0	8	15.4
Don't Know	4	10.8	2	13.3	6	11.5
Total	37	100	15	100	52	100

of water, water is capable of fulfilling our premises most completely. Hence, water would be our most valid conclusion and such responses as snow or ice and some sort of food would be among the preferred responses. This problem involves drawing implications from a set of premises about *something* rather than *someone*. The premises must be related by eliminating assumptions violating any one of the six properties ascribed to the subject.

The preferred responses occur systematically more frequently in Group A than in Group B. The nonpreferred responses (mud or sand and objects) reverse this trend, Group B producing twice as many nonpreferred responses as does Group A.

It is valuable to determine the number of subjects in each group offering divided or furcated responses. Such an analysis is made in Table 16. We note the significantly greater number of Ss in Group B who produce furcated responses. Furcation has the net effect of leading the S away from his central task: the employment of all of the information given him. Hence, the preponderance of nonpreferred responses occurs in Group B. Twelve of the fifteen participating Ss in Group B produced one of the more preferred responses, six of these being the most adequate response: water. In the A group, all but three Ss produced at least one of the preferred conclusions. The problem was apparently a comparatively simple one for most of the Ss.

A final qualitative finding is of special interest to this study. We examined the response sequences to item 6 for each group to attempt to reconstruct the reasoning that led to the particular conclusion offered by the child. A striking difference was found between the sequence given by members of Group A and that

by members of Group B: 8 of the 15 children in Group B divided their responses into segments, each segment fulfilling one of the six qualities mentioned in the statement of the problem; only 2 of the 37 Ss in Group A demonstrate such segmentation. An example of this reasoning follows:

F3: "Maybe a food you can eat. A deep freeze is so heavy. The soft is some orange juice, the hard is a deep freeze. A finger you can't cut off because it is a bone. A ball is hard you can throw it. A sandwich you can eat and freeze and jump into."

S10: "Mud you can slop in it, and the other one is a ball, it's hard. Water you can turn into ice, slip in it."

Typically, not all of the six qualities are mentioned. *S* asks himself: "What is hard?" and "What can I fall into?" and "What can you boil?" It seems as if the subject cannot conceive of something with multiple attributes or is trying to start an association beginning with one determinant (i.e., such as hardness or swallowability). The process of *furcation* described above differentiates between the A and B groups beyond the .01 level of probability.

Evaluating the evidence presented in this chapter, we conclude that in situations requiring deductive reasoning, more prejudiced children tend to produce no more numerous but more invalid conclusions than do less prejudiced children. The sheer number of responses does not differentiate Group A from Group B as it did in our test of inductive reasoning. Moreover, if all Ss' responses are pooled and the two groups are compared for the frequency of valid and invalid conclusions, again there is no statistically significant difference. We may divide all the problems into two types: those that are syllogistic and those that are nonsyllogistic in form. We find no differences between Group A and Group

B responses to the syllogistic problems. Prejudiced children, however, have much more difficulty in arriving at valid conclusions when the problem and/or its conclusion is not in syllogistic form. Unprejudiced children have significantly less difficulty. Prejudiced children, when they do arrive at a valid conclusion, do not generally tend to "see" the conclusion at once but apparently require one or two "trial balloons" before reaching it. A statistically greater number of the unprejudiced children arrive at the correct conclusion at once, satisfying the restrictions of the premises. Prejudiced children tend to *furcate* the ambiguous problems far more than do unprejudiced children. *Furcation* tends to concretize and simplify the problem, essentially transforming it into a problem of "formal logic." Similarly, more prejudiced children tend to convert ambiguous problems into unidirectional problems (i.e., as in item 4). Unprejudiced children engage in *furcation* and conversions of the sort described far less frequently.

Finally, prejudiced children tend not to admit a lack of knowledge but persist in offering hypotheses, even when they are challenged by *E* concerning their validity. Over-all, there are 52 "Don't Know" or "I give up" responses in Group A, an average of about 1.3 such responses per *S* to the 8 test items of the DRT. In Group B there are 10 "Don't Know" or "I give up" responses or .56 responses per *S*. This difference between Groups A and B on the tendency to admit lack of knowledge or an inability to reach a solution is significant beyond the .01 level of confidence. It would appear, then, that prejudiced children are not only less capable of producing valid conclusions but the invalid ones they *do* produce are *dogmatically held*. Another way of interpreting this result is that

prejudiced children tend not to admit what, to *E*, might be seen as an admission of weakness. The validity of this hypothesis would seem to receive support from the "anti-weakness" tendencies of the ethnically prejudiced found in the *California Study* (1). (See also 47, 48, and 49.)

PREJUDICE AND PROBLEM SOLUTION

Rokeach (44), in a brilliant research carried out as part of the *California Study*, demonstrated the existence of a "generalized mental rigidity" in ethnocentric adults. He found similar results in the part of his research dealing with late childhood and adolescence (the median age of his group was 13 years). Since this latter aspect of his work is of immediate importance to the present study, we shall report it in detail.

Rokeach administered the *California Ethnocentrism Scale I* to 193 subjects and followed this with a test of mental set: the subject was asked to solve a series of arithmetic problems involving the use of three hypothetical jugs of given capacity. Set was established if *S*, after solving three problems by a long and devious method employing all three jugs, continued to employ this "less adaptive" technique over a group of "critical" problems (five in number) in which the problem was capable of being solved by a short, direct method as well as by the long, indirect method. Only subjects falling into the set and who had solved control problems by the short method were considered in the final analysis of the results. Of the original 193 subjects, 75 comprised the final group which was divided into those High and those Low in ethnocentrism as measured by the *California Scale*. Rokeach found that the High children gave a greater proportion of "rigid" solutions (i.e., using the long method) than did the Low children on all of the 10 critical problems employed. Statistical significance differentiating the two groups was found on 4 of the 10 problems. Analyzing the amount of scratch paper used by each group as a measure of rigidity (Rokeach sees the scratch paper as a "mental crutch," unnecessary when *S* uses the short method), it was found that the High group used more scratch paper than did the Low group in 12 of the 13 problems. The differences between the

two groups were found to be significant on only one problem, however. Rokeach concludes that those who are rigid on tests of social attitudes (ethnocentrism in this case) are also rigid in solving nonsocial problems.

Both clinical and experimental studies (43, 53, 58) have tended to establish the general nature of rigidity: the tendency to fixate responses, to engage in stereotyped modes of behavior and thought, to persevere ideationally, and to think and behave in concrete as opposed to abstract ways. Rigidity in the present research is taken to mean the inability to solve conceptual problems due to the perseveration of a previously established mental set, no longer adequate to their solution. If, indeed, prejudice may be thought of as a form of perseveration either of thought or behavior or both, and if perseveration is a general factor of personality organization, then it should be found in cognitional processes as well (10).

It may be argued that children do not possess deep-seated prejudices: that, in fact, children at this age do not exhibit rigidities in any systematic fashion. Our findings concerning the prejudiced seven-year-old child's ability to think and reason logically would not tend to support this view. Rather, it would appear that more prejudiced young children do exhibit a narrowing of mental capacities appearing in the form of simplification, furcation, and concretization in solving various types of problems requiring different modes of approach.

In this chapter we are concerned with the ability of *Ss* to solve certain conceptual problems involving the establishment and breaking of induced mental sets. While mental sets are not to be regarded necessarily as detrimental to adequate problem solving, in the present instance they were especially arranged so as to have this effect. Our problem stems from a common observation: that the dogmatically prejudiced person tends to be "narrow-minded,"¹⁷ unamenable to persuasive argument. We may say that his

¹⁷ Rokeach has extensively explored this concept (see 45, 46).

views appear to be unusually fixed or set and unusually resistant to change. One may therefore argue that prejudice is significantly associated with the general characteristic of inflexible mental sets: that mental sets are to be found among highly prejudiced persons while this is less likely to be the case with individuals showing little prejudice. Stated formally, we shall attempt to test two hypotheses, each with two subparts: (a) more prejudiced children tend (i) to establish *inefficient* mental sets, and (ii) to employ inappropriate methods to break them; and (b) less prejudiced children tend (i) not to establish *inefficient* mental sets, and (ii) to employ appropriate methods to break them when they are established.

Our procedure was designed to have *S* search for a concept that would allow him to predict the location, the position, and the sequential movement of an object hidden among a group of similar objects. Four separate subtests comprise the total test: a location test, a position test, a movement test and a reverse-movement test.

The materials employed were a group of three white, flat, push-drawer pillboxes measuring $2\frac{1}{2} \times 1\frac{1}{2} \times \frac{3}{8}$ inches. The boxes were of uniform appearance and were presented to *S* with seams facing the same direction. Since standard pillboxes were employed they were easily replaced as soon as they became worn or defaced. The *S* was seated and *E* began:

"Today we have a guessing game. Do you like guessing games? Good. Now listen very carefully. Do you see those three boxes?" (*E* places the boxes in front of *S* and points to each in turn). "You see, here are three boxes that are all alike. But, on the back of each box there are little dots. See, this one has 2 dots, this one has 3 dots and this one has 4 dots. Now here's how we play this game. I'll mix the boxes while you hide your eyes and then you guess where one of the boxes is. We'll look for the same box until we find it 5 times in a row. If you guess right I'll say 'Right' and show you the bottom of the box. If you guess wrong, I'll say 'wrong' and show you the right box. I'll never try to fool you, but I'll show you the right box each time. Each time you guess there will be a trick to the puzzle. If you guess the trick, you'll know where the right box is. All right, do you understand the game?" (If *S* indicates understanding, *E* proceeds to the trial

test. If *S* is uncertain or does not understand *E* re-explains the procedure and rules until *S* indicates understanding. In several instances the trial test was held to familiarize *S* with the procedure. As a result, no *S* was eliminated from the testing because of lack of comprehension of the nature of the "game.") "All right, now hide your eyes until I tell you to look."

A trial test was held to prepare *S* for the actual testing procedure. The three boxes were laid before *S* in a vertical-horizontal-vertical sequence. He was told to "look for box No. 2" (with 2 dots on the bottom). The correct box was kept in the horizontal position. The *E* permitted *S* to "guess" the correct box until he made 3 consecutive correct choices and verbalized the solution. The verbalization consisted in *S* noting that box No. 2 remained in the same relative position even though its location changed from trial to trial. (*E* explicitly pointed out that the "trick" was horizontally. The other two boxes were kept in the vertical position). If *S* did not correctly identify the sought-after box within 10 trials, *E* described the solution for him. The *S* was then presented with a second test trial with a horizontal-vertical-horizontal (H-V-H) sequence and the same procedure was followed as with the first test trial. No subject failed the second test trial. The *E* immediately proceeded to main variations of the test.

Test 1 was a *location* test. The *E* was presented with a Left Oblique-Vertical-Horizontal Pattern. The key box changed *position* randomly to vertical, horizontal, and oblique, but remained as the left-hand box (facing *E*, right-hand facing *S*; locations are given with the *E* as point of reference throughout this test). *The S was allowed 25 trials to locate the correct box. Five successive correct trials and the verbalization of the principle were the criteria of success.* If *S* reached the 21st trial without a chance of reaching the criterion of sequential successful trials, the test was discontinued. Successful completion of Test 1 was prerequisite to being considered in the analysis of results, since this test served as the "set-inducing" stimulus. Nevertheless, for the sake of uniformity all subjects were administered all four subtests (with certain exceptions noted below). If *S* verbalized the principle before the 5 successive correct trials, the solution was considered achieved. Usually, if *S* achieved a solution within the first 5 trials, *E* would press *S* to go on with the test in order to determine whether the "solution" allowed prediction to further successful trials.

Test 2 consisted of a *position* test. The correct position was the Oblique with the *S* again being required to choose between vertical, horizontal, and oblique position. In this case the *location* of the objects was changed randomly at each

trial. The same criteria used in Test 1 for success were also employed on Test 2. On this test and the two succeeding tests the *S* recorded the *S*'s choices, and all remarks made by *S* verbatim, on a specially prepared scoring sheet. The sheet was arranged in the form of a contingency table with the random locations and positions marked across the X-axis and *S*'s selections marked along the Y-axis. For Test 1, locations were marked along the Y-axis and positions marked along the X-axis.

Test 3 was a *movement* test. The boxes were presented to *S* in a V-V-V sequence. The correct box was located at the *left-hand side* at the start of the test and was shifted successively to middle, then right, then left, middle, right, left, etc. until the criterion was reached or *S* failed.

Test 4 was a *reverse-movement* test. The boxes were arranged again in a V-V-V pattern with the correct box being placed on the *right-hand side* and shifted successively to middle, left, right, middle, left, right, etc., until the criterion was reached or *S* failed.

The results are based upon 53 *Ss*, 38 (72%) of whom are in Group A, and 15 (28%) of whom are in Group B. These figures are only slightly different than the ratio of 70%/30% for the entire group.

We assumed that most *Ss* would pass Test 1 since it consists of a simple localization problem. Only three *Ss* failed Test 1 and they were dropped from the analysis of the results. Hence, 53 *Ss* passed Test 1. We further assumed that Test 1 would serve as a set-establishing problem: *S* would be set for location-solutions when Test 2, requiring the subject to drop the first set and seek for a position-solution, was administered. If no set was established by Test 1, *S* would have little trouble in arriving at the solution of Test 2. As shown in Table 17, 32 of the 53 *Ss* failed Test 2 (60.4% of the group).

Test 3 offered *S* an opportunity to break the original set, if it had been established, since it removes one of the characteristics of the two previous patterns: all boxes are presented in a V-V-V pattern. The subject does not have to be concerned with the position of the boxes. Moreover, since Test 3 is essentially a

TABLE 17
SUCCESS AND FAILURE FOR TOTAL GROUP
ON TESTS 1-4: PROBLEM-SOLVING

Test	Pass		Fail		Total	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
1	53	100	0	0.0	53	100
2	21	39.6	32	60.4	53	100
3	39	73.6	14	26.4	53	100
4	27	50.9	26	49.1	53	100
Totals	140	66.0	72	34.0	212	100

location problem, the *S* can draw upon his skill and success in achieving a solution on Test 1 in attempting to solve Test 3. We note in Table 17 that only 14 *Ss* failed Test 3 (26.4% of the total group).

Test 4 was given to determine whether or not the original set was still operating in those *Ss* who failed it and to assess the effect of Test 3 as a set-inducing stimulus for those *Ss* who passed it. Again we see that 26 *Ss* failed Test 4 (49.6% of the group).

How many tests were passed and failed by our two groups? Do prejudiced children establish inappropriate perseverative mental sets in greater number than do unprejudiced children? We analyzed the frequency of failures for each subtest according to prejudice groups. Table 18 summarizes our findings. Of 152 individual tests taken by Group A, 113 (74.3%) were passed and 39 (25.7%) were failed. Of the 60 individual tests taken by Group B, 27 (45.0%) were passed and 33 (55.0%) were failed. This difference is reliable: Chi square is 16.43 and p is $< .001$. If we eliminate Test 1 data from the results (since all *Ss* passed Test 1) the differences are still great: Chi square is 19.75 and p is $< .001$.

We conclude that the more prejudiced children are less able to solve our critical problems than are the less prejudiced

TABLE 18
SUCCESSSES AND FAILURES OF GROUPS A AND B ON TESTS 1-4: PROBLEM SOLVING

SUCCESSSES AND FAILURES OF GROUPS A AND B ON TESTS 1, 2, 3, AND 4									
Test	Group A				Group B				Total
	Pass		Fail		Pass		Fail		
	N	%	N	%	N	%	N	%	
1	38	100	0	0.0	15	100	0	0.0	53
2	18	47.3	20	52.7	3	20	12	80.0	53
3	32	84.2	6	15.8	7	46.7	8	53.3	53
4	25	65.8	13	34.2	2	13.3	13	86.7	53
Totals	113	74.3	39	25.7	27	45.0	33	55.0	212

children. We find (Table 18) that the ratio passing to failing of Groups A and B on Test 2 is not statistically different. We note that while 53% of the Ss in Group A failed Test 2, 80% in Group B failed it. Similar ratios between groups A and B on Test 3 are significant at the .02 level. For Test 4 the comparable p is $< .001$. Hence, there are increasingly significant differences between the pass-to-fail ratio of Group A as compared to that of Group B.

We turn our attention now to the pattern of successes and failures in each group. Which tests were passed and failed in sequence by our two groups of subjects? We find that 15 Ss passed all 4 tests and all but one are in Group A. Of the 14 Ss who passed 3 of the 4 tests, 13 are in Group A and 1 is in Group B.

Hence 26 of 28 Ss passing at least 3 tests are in Group A. Seven members of each group passed only 2 tests and 4 members of Group A compared to 6 members of Group B passed only one test. This data is summarized in Table 19.

If we note the number of Ss in each group who, having failed Test 2, went on to fail Tests 3 and 4, we find that Group A differs from Group B in this regard with significantly greater frequency of failure in Group B. Thus, while only some members of Group A failed on Test 2, about the same number went on to fail the other tests as well. However, in Group B, children who failed Test 2 generally failed either or both the remaining tests.

If we take the frequency of failure to solve more than one of the "critical" tests

TABLE 19
PATTERNS OF SUCCESSSES AND FAILURES ON TESTS 1-4: PROBLEM-SOLVING

Tests					Group A		Group B		Total	
					N	%	N	%	N	%
1	2	3	4							
*	*	*	*		14	36.9	1	6.7	15	28.3
*	X	*	*		11	28.8	1	6.7	12	22.6
*	*	*			2	5.3	0	0.0	2	3.8
*	*	*	X		2	5.3	1	6.7	3	5.7
*	*	X	X		2	5.3	6	40.0	11	20.8
*	X	*	X		5	13.2	6	40.0	10	18.8
*	X	X	X		4	10.5				
Totals					38	100	15	100	53	100

Legend: * Pass
X Fail

(i.e., Tests 2, 3, and 4) we find that Group A passes 27 and fails 11 tests, while Group B passes 2 and fails 13 tests, and this difference is reliable. It seems clear that the more prejudiced children fail a good many more tests than do the less prejudiced children. From Table 19 we find that 47% of the Ss in Group B fail two "critical" tests, while an additional 40% fail all three such tests. Comparable percentages for Group A find only 18.6% failing two tests and 5.5% failing all three.

Twenty-four Ss failed two or three tests. Of this group only 3 passed the first 2 tests and failed Tests 3 and 4. The remaining 21 Ss failed either Test 2 and 4 or Tests 2, 3, and 4.

We now turn to our central problem—do prejudiced children, in fact, tend to establish perseverative mental sets in greater number than do unprejudiced children? Our task was to compare the frequency of success on each of Tests 2, 3, and 4 and the frequency of failure due specifically to a perseverative mental set for each group. We first set up a group of criterion categories to establish which responses could be attributed to set and which responses accrued in consequence of other factors. Our general procedure was to analyze the sequence of "guesses" made by each S who failed Tests 2, 3, or 4. In Test 2, for example, we wanted to know what procedures S followed, how systematic he was, and whether he attempted to set up "hypotheses" concerning the nature of the "principle" that would lead to success. Test 2 involved a position set (the oblique box is the correct object). Did the S persist in searching for a correct location, the principle of which had led to success on the first test? How random are the failing Ss attempts to solve the problem and "get the trick?" In brief, how and why did the

S fail? The criterion categories established were as follows:

1. *"Location-set."* Three categories (S selects either left, middle, or right location at least twice as often as the other two locations combined).

2. *Avoidance.* The S selects one of the locations less than one-half as frequently as the location chosen most frequently.

3. *Faulty prediction.* The S correctly identifies the object being sought but cannot verbalize the principle or predicts the movement or change of position incorrectly.

4. *Wrong position.* The S incorrectly chooses a position, other than the oblique, on Test 2 and persists with this selection.

5. *"Superstition."* The S believes that a given position "belongs" in a given location or that some position "should not belong" in a given location.

6. *Arbitrary hypothesis.* The S makes an unwarranted assumption about the nature of the solution and persists in this assumption over a number of trials.

7. *Cue-spotting.* The S picks out some real or imagined surface flaw on the box and follows the boxes with his eye to detect the flaw before making his selection. This procedure did not continue very long, since after five successive correct choices E asked S for the principle. The S would usually either admit the cue-spotting as his solution or say that he hadn't any idea. In either case E performed a quick sleight-of-hand in substituting a fresh set of boxes.¹⁸

8. *Random selection.* The S makes no attempt to locate the correct box but either closes his eyes and swings his hand over a box to touch it or wildly aims at a box or merely guesses on each trial without attempting to work out a principle.

9. *Movement pattern.* The S apparently realizes that the correct box does not remain in the same location on every trial. He, therefore, tries out several "hypotheses" as to the correct location. He assumes that the location of the correct object changes from trial to trial. Attention is not paid to the position of the object. A location-set may yet be operating but no single location is given preference over others.

The above nine categories are not mutually exclusive, since it is possible for a child to exhibit several types of response during the session. Nevertheless, most of

¹⁸ Cue-spotting was also defined as S's attempt to pick the right box by watching E's facial expression or the direction in which his pencil was pointing, etc., as S hovered over the boxes. One child stopped at each box, looked at E and moved to the next until he made a choice.

our Ss tended to be consistent in falling into only one or two response categories. The findings reveal that there are a few differences between Groups A and B in the manner of approach to the set problems that lead to failure. *There are few cases of perseveration of set in either group* (i.e., left-location would be evidence of such a carry-over from Test 1). Both groups show equivalent tendencies to interpret the instructions literally. We gauge this tendency from the existence of "middle-set" and "right-set" responses. Apparently Ss making such responses are responding to the mandate to "find another trick" by searching for a "location-trick" but in a *new* (other than left) location. Hence, we may say that some aspect of the location-set and right-set responses persisted. The response of "left avoidance" would also indicate a concrete, instruction-bound type of response. The child, assuming that he must find a "new" location in the next test, avoids the left-location because "left" was the correct solution of the first test. Hence, he divides his responses about evenly between middle and right. (All four of these categories—left, right, and middle sets and left-avoidance) are essentially *location* sets: S is not responding to the position of the objects in as systematic a fashion as he regards their locations with respect to one another. Hence, all four of these categories give evidence of perseveration from Test 1.

The "random selection" category provides the first significant qualitative difference between Groups A and B. Prejudiced children, more frequently than non-prejudiced children, tend to choose at random from trial to trial, rather than trying to seek some systematic relationships between their successful and unsuccessful choices. They guess the response. Some children do this by swinging their

arms over the boxes in picking an object—any object—they may fancy. Others try to devise a scheme of some kind (i.e., eenie, meenie, minie, mo, catch a tiger by the toe, etc., or 10-20-30-40, etc., to 100), in the hope that the final word or digit will fall on the correct object. Random selecting apparently increases in frequency from test to test. Nine Ss (3 in Group A and 6 in Group B) exhibited this response on Test 4, although a total of only five such responses are found in Groups A and B combined on the two previous tests.

Cue-spotting occurred as frequently in Group A as it did in Group B. In a sense, cue-spotting is about as random as is random-selecting. The difference between the two is the systematic fashion in which the S watches the boxes or E for some sign of the correct box during cue-spotting, as compared to the desultory "calling" in random-selection.

The category "faulty prediction" yields equivalent frequencies for each group. The S has the idea of position (in Test 2) and of movement (in Tests 3 and 4) and may even select the correct box for more than the required five consecutive trials but incorrectly verbalizes or predicts the position or type of movement necessary to achieve a solution. It is possible that some of these Ss actually have the solution but simply cannot express it adequately. They were scored as having failed the test, however, since it was decided to allow a successful response only when both the criteria of consecutive correct trials and prediction or verbalization were met.

Arbitrary hypotheses are offered in six cases by Group B and only one case by Group A. The six responses in Group B represent 13.5% of the total responses for Group B—the second greatest percentage of responses in that group. The

six arbitrary hypotheses offered in Group B are the work of five subjects. They occur on tests 3 and 4 indicating the increasing randomness of response with increasing failure in some subjects.

The category "wrong position" is exceedingly interesting. Only three subjects gave this response, all in Group A and all on Test 2. Essentially, this is a set-breaking response. The *S* assumes that he must now look for another principle but mistakenly selects either of the two positions that are not correct. This is probably a form of rigidity since *S* apparently "wants" the correct position to be the one he selects. Nevertheless, he does recognize that he is not seeking a particular location. His persistence with a position that is consistently incorrect is, however, just as maladaptive to the situation as any other form of error mentioned above.

The "superstitious" response appears with equal frequency in Groups A and B. The *S* persists in the belief that one or more positions "belong" in given locations. Therefore, he will not, for example, select the oblique box in the left-hand location, but will pick it only when it appears in the middle location. This type of response is probably a location-set but is coupled with position beliefs probably because they have a disturbing influence on *S* who is trying to concentrate on locations.

The final category is the movement pattern response. This is the most adequate type of response that characterizes the errors made by *Ss* in the course of failing to solve the problems. Unlike the location-set responses mentioned above, here the *S* attempts several "provisional tries" (to borrow another term)—all within the framework of finding the correct location by establishing the movement pattern for the correct box. The usual

response is that: "It moves to different places." The "it" in question is not the position of the objects (i.e., vertical or oblique boxes do not move) but simply the correct object.¹⁹ The difference in frequency between Groups A and B on this category is short of significance.

The above analysis concerned itself with the *responses* of subjects on our problem-solving tests. We now proceed to test our hypotheses concerning the operation of perseverative mental sets in our two groups of subjects. Since a subject may produce more than one response on each test, each subject who failed a test was placed into one of three categories: (a) location or movement set, (b) adequate methods, and (c) inadequate methods. For the most part, location sets were found operating on Tests 2 and 3, while a movement set was found on Test 4. We have already described a location set. A movement set resulted when an *S* attempted to use the same formula on Test 4 that had operated on Test 3 (i.e., left, middle, right, left, middle, right, etc.). An "adequate" method was described by the *S*'s realization of one of the essential qualities of the solution (i.e., the movement of the objects rather than their stationary character, typical of the set solution in Test 2). The "faulty prediction," "wrong position," and "movement pattern" responses were thus defined as adequate responses. "Inadequate" method was defined as a failure to deal with the essential problem of each test in an other-than-set fashion. Thus, the "random selection," "cue-spotting," "arbitrary hypotheses," and

¹⁹ The large number of random selection responses on Test 4 may in part be attributed to *S*'s failure to be aware of the *movement* of the objects in systematic fashion. Therefore, *S* gives up his attempts to work out a solution. To him, the movement is arbitrary and random.

"superstition" categories are considered as inadequate responses. In all, there are 72 cases of failure, 39 in Group A and 33 in Group B. *It has already been demonstrated that the pass/fail ratios of the two groups are significantly different.*

Do prejudiced children fail the critical problems more frequently—and fail because of the interference of perseverative sets?

From Table 18 we noted that 18 subjects in Group A passed Test 2 and that 20 failed. Of these 20 failures, 5 failed because of perseveration of the principle of Test 1. Hence, of the 38 unprejudiced subjects, 13.2% failed Test 2 due to a perseverative mental set. In Group B, 12 of the 15 subjects failed Test 2. Of these 12 failures, 4 or 26.7% were due to the perseveration of the principle of Test 1. These differences are not significant.

Turning to Test 3 we find that, in Group A, 32 subjects passed and 6 failed; and, of these latter, 2 fell into a perseverative set. In Group B, 7 subjects passed and 8 failed Test 3. Only 1 subject failed due to a set. Again, the group differences are not significant. On Test 4, 4 subjects, of the 13 failing in Group A, failed due to set. There were no set failures in Group B on Test 4.

Obviously, from these results we must conclude that pernicious mental sets were *not* the primary cause of failure on these tests. Even if they were, we have not been able to detect them. It remains for further research to determine whether this conclusion is reliable, especially in view of Rokeach's conflicting results. It should be re-emphasized that we are here dealing with *failures* on tests of problem solving. Great differences have been found between our two groups in their respective ability to *solve* these Problems.

Slightly more than half of the 39 fail-

ures in Group A were problems whose solutions were not far from accuracy. Only 21% of the failure solutions are "inadequate." In Group B on the other hand the comparable percentages are 27% and 58%. These differences are reliable at the .01 level of significance.

These rather unexpected findings lead us to suspect that gross cognitive differences *do* distinguish prejudiced from unprejudiced children. Major evidence in support of this view comes from the fact that the general method of approach to the problems appears to be quite different.²⁰ Tentatively, we may state that:

1. Unprejudiced children appear to be more task-oriented to the problems, staying within the confines of the demands of the situation. This is indicated by the significantly greater number of "adequate method" failures, as well as by the high rate of solutions, as compared with the prejudiced group.

2. Prejudiced children tend to become discouraged and perplexed rather easily over the problems. Perhaps we are here dealing with the effects of frustration induced by failure. We feel that such motivational factors as low frustration tolerance and low threshold for feelings of defeat are aspects of what might be called a "prejudiced personality pattern," closely related to cognitive functioning. Such categories of response as "random selection" and "arbitrary hypothesis" are the cognitive counterparts of feelings of despair, intolerance for intellectual tasks, and a need to do *something* that approximates the demands of the problem.

3. From the point of view of insight or understanding of the problem and its solution, prejudiced children are vastly inferior to unprejudiced children

²⁰ As indicated below, intelligence appears to play a role in these differences, if an obscure one.

on this test. Not only do children in Group B fail more frequently than Group A children, but in failure they are significantly further from any acceptable form of solution after the standard units and time limit than are Group A children. Recalling that general intelligence for each of the groups is equivalent, we feel justified in assuming that the differences are more likely functions of "personality," attitudinal and motivational factors than they are of special intellectual factors.²¹ This is not meant to deny the importance of the organization of intelligence in individual children or to set forth a dualistic "mental-faculties" argument. We merely wish to point out the fact that the nature of our results requires a far broader interpretation than one solely encompassing the cognitive functions.

²¹ The argument may be raised that the incidence of set-induced failures on our test of problem solving may be attributed to a deficiency in intelligence. In this instance such an argument finds support. The mean IQ of those who passed all four tests is 118.0 ± 10.8 ; the mean IQ of those who failed two or three tests as a consequence of some form of inefficient mental set is 111.1 ± 10.8 . The difference between these two means is significant at the .05 level.

Breaking the data down into component subgroups, we find that no differences exist between the mean IQ's of those who pass all four tests and those who fail two or more tests in Group A (mean IQ of the latter group is 114.9 ± 7.4). Since only one subject passed all four tests in Group B, no similar comparison may be made. The major contribution to the depression of the mean IQ among those who failed is made by 12 subjects in Group B. Their mean IQ is 101.1 ± 10.7 , differing significantly from those in Group A who pass all tests, and approaching but not reaching significance in the mean IQ difference between those who fail in each group.

Hence, intelligence appears here to be associated with problem-solving ability: The failing subjects are generally less intelligent, while the more intelligent subjects find solutions. The meaning of these results is not altogether clear, since we do not know what function problem-solving ability plays in determining the final IQ.

Summing up, we set out to test the abilities of our two groups to solve certain conceptual problems. Our hypothesis centered on the point that the more prejudiced children would more likely fall into inappropriate mental sets and hence fail the problems more frequently than would the less prejudiced children.

We found that while children in Group B failed our critical tests more frequently than did children in Group A, the former did *not* fail due to the presence of a perseverative mental set any more frequently than did the latter. In examining the determinants of failure we found that children having anti-ethnic attitudes tended to employ methods highly inappropriate to the solution of the problems. Children possessing few anti-ethnic attitudes tended to come closer to a solution making "good" errors even when they failed to reach a correct solution.

It was speculated, with some evidence, that there *are* important cognitive differences between our two groups. Although perseverative set was not prominent as a factor, such items as low frustration tolerance, "anti-intellectualism," lack of insight, and a lack of task orientation were important factors determining our results.

Inasmuch as our results tend to conflict with those of Rokeach and as a consequence of the actual differences obtained between the performance of our two groups, additional research in the area of problem-solving behavior and ethnic attitudes seems warranted.

PATTERNS OF MENTAL FUNCTIONING— SUMMARY AND INTERPRETATION

In previous sections we have studied the thinking of ethnically prejudiced and unprejudiced children. In the present section we shall attempt to bring to-

gether our findings and to determine what "typical" patterns of mental functioning seem to characterize the prejudiced and unprejudiced child. We shall also venture some tentative hypotheses concerning the role that such patterns play in the total personality organization of each type of child. Finally, we shall attempt to draw some general implications from the present research.

Patterns of Mental Functioning

The central purpose of the research reported in this paper was to determine whether certain special types or qualities of mental functioning distinguished children who were more and less ethnically prejudiced. Examining the results obtained from our five tests, certain clear tendencies emerge, which, by their direction and statistical significance allow us to depict typical patterns for each of our two groups of Ss. Hence, what follows must be regarded as broad approximations of the thought processes of typical prejudiced and unprejudiced children. Four "areas" of mental functioning will be discussed in turn for each group. These are ethnic attitudes, generalization (concept formation), deduction (drawing implications from given facts), and mental flexibility.

1. *Ethnic attitudes.* Sixty seven-year-old children of better than average intelligence were given five tests. On the basis of an Ethnic Attitude Test each child was categorized as "more" or "less" prejudiced. The Ethnic Attitude Test was composed of items designed to elicit statements agreeing or disagreeing with strongly prejudicial assertions attributed to "other persons." Forty-two unprejudiced and 18 prejudiced children were found to fit the criteria established to evaluate responses to the Ethnic Attitude Test.

Analysis of the manner in which the prejudiced child responded to various ethnic groups revealed a tendency to generalize, categorize, and dichotomize. Ethnic outgroups were regarded as universally without virtues of any sort, while ethnic ingroups were exaggeratedly praised and

glorified. The ethnically prejudiced child tended to see other groups and their members in stereotyped or clichéd roles and as violators of rules of conduct to which he meticulously adhered. He thought of both ingroups and outgroups as fitting two ends of an all-or-none moral dichotomy: good or bad, dirty or clean, kind or unkind, etc. He tended to overgeneralize his responses on the basis of a single instance: he will feel hostile to Negroes because he heard that one Negro was not honest.

By contrast, the unprejudiced child tended to think of ethnic groups in individualistic fashion. He rejected sweeping generalizations made about his own group as well as about other groups. He tended to deny unjustified glorification of the ingroup. He qualified categorical statements with provisional or conditional statements (e.g., "Some colored might be dirty, sometimes"). He tended to equalitarian sentiments: other groups are just as good, smart, etc., as his own. He may deny any differences between other ethnic groups in countering attacks against outgroups.

A high degree of outgroup acceptances was coupled with a rejection of ethnocentrism. Denial was made of statements which attributed superlative virtues to Americans. Equal status, equal ability, equal potentiality was the central theme of his attitude toward each ethnic group.

2. *Generalization (concept formation).* A child was confronted with a concept formation test consisting of eight odd-colored and odd-shaped paper objects. He was required to find the "trick" that would yield equal groups of the identical "kind" of object. The concept of size would solve the problem, since only two sizes were presented. The concepts of color and shape yielded unequal groups.

An analysis of the performance of the prejudiced child on this test revealed:

1. A high incidence of failure to form the appropriate concept.
2. A tendency to become "stimulus-bound": to be incapable of divorcing the impelling surface perceptual attributes of color and shape from the situation in order for the relational concept of size to be admitted.
3. A tendency toward concretization as opposed to abstraction in approach to the objects: a high incidence of "personification" of the objects as parts of a jigsaw or as segments of some larger object.
4. A tendency to jump to the conclusion without testing "hypotheses" concerning the ultimate solution.
5. A tendency to persevere an hypothesis in spite of previous vain attempts to employ it toward achieving a solution.

6. Incapacity for making use of aids from *E* toward the solution.

Performance of the unprejudiced child on this test was marked by:

1. High incidence of success in achieving the appropriate concept either spontaneously or with aids from *E*.
2. Capacity for appropriately adjusting his attempts toward solving the problem in terms of aids given by *E*.
3. A markedly high incidence of "adequate" or appropriate "hypotheses" toward the solution of the problem.
4. A high incidence of "personification" and perseverative responses that nevertheless culminated in the appropriate solution.
5. Marked capacity for ignoring the irrelevant surface attributes of the objections and for dealing with the objects on an "abstract-conceptual" plane.
6. A tendency to ruminate or "mull" over the possible solutions before offering them as "final."
7. A tendency to offer *alternate* hypotheses when those attempted prove fruitless: to suggest a variety of potential solutions.

3. *Deduction (drawing implications from given facts)*. In general, prejudiced children tend to find little difficulty in solving simple, strictly logical, formalistic, linear, deductive problems. However, in attempting to solve deductive problems in which alternate solutions are possible or where the problem involves the relationship of a complex of facts, the prejudiced child has difficulty. Typically, he attempts to *alter the nature of the problem* from one of comparative ambiguity to one of clear formal logic. In so doing he tends to *simplify* the problem by *furcating* it. He then proceeds to offer a "solution" for each furcated segment. The problem is seen as a group of separate problems, each with an independent solution. He tends to reconstruct them by removing the ambiguity: only one reply is possible. Further, the prejudiced child feels *certain* that his solutions to such problems are the only ones possible. Even in failing to see the errors he commits, he *dogmatically* holds to the solutions offered. Because they are oversimplified, his solutions tend to be relatively *concrete*, that is, they are determined by one premise, isolated from the complex of premises making up the problem.

The unprejudiced child has more difficulty with ambiguous than with formal logic problems but has significantly less difficulty than does the prejudiced child. He tends to solve the ambiguous problems in terms of their complexity, to relate the various aspects of the problem so that the conclusion drawn is based upon the total group of attributes assigned to the premises.

He tends neither to simplify nor to furcate the problems, nor to convert the ambiguous problems into formal logic problems. He may offer equivocal replies or specifically point out that the problem may have different answers. He tends to be undogmatic in his final solutions. When challenged he will admit the potentiality of alternate solutions, if he has not already offered them spontaneously.

4. *Mental flexibility*. All subjects were offered a test consisting of four subparts. In each instance, the child was presented with three rectangular boxes. On Subtest 1, the child was to "locate" the correct box which was rotated in three positions although remaining in the same place relative to the remaining boxes. The latter were similarly rotated in place. Subtest 2 made a given position (oblique) "correct" while all boxes were shifted to different locations on each guessing trial. On subtest 3, the child was confronted with three boxes all standing in the vertical plane. He was to find the box that began at the left hand and was sequentially moved by step intervals (to middle and right and left and middle, etc.) on each trial. Subtest 4 reversed the order of movement, the correct box being located at the right on the first trial.

Fifty-three Ss passed subtest 1, the criterion for being included in subsequent analysis of the results. An important source of failure was the high incidence of *maladaptive* or "*inadequate*" methods applied to the critical problems. This latter cause of failure appeared to be in part a function of *low frustration* tolerance among the prejudiced group. Excessive *random guessing* and the use of *arbitrary concepts* characterized the "*inadequate*" attempts to solve the problems.

The unprejudiced child tends to offer hypotheses concerning the solution that are in the proper direction: they are adequate to the solution of such problems as are offered them although they fail to find the precise solution. These latter are "good" failures.

Bringing the types of mental functions characterizing prejudiced and unprejudiced children in our study into sharp relief, we find the following patterns:

Prejudiced Child

1. rigidity
2. overgeneralization
3. categorizing and dichotomizing
4. concretization
5. simplification
6. furcation
7. dogmatism (omniscience)
8. intolerance of ambiguity

Unprejudiced Child

1. flexibility
2. realistic generalization
3. individualizing
4. abstraction
5. retention of complexity
6. retention of totality
7. lack of dogmatism
8. tolerance of ambiguity

A closer examination of the above patterns reveals that two major characteristics of mental functioning typify the prejudiced child. These are mental rigidity and intolerance of ambiguity. The less prejudiced child concedes his own weaknesses and seems more interested in what the task demands of him than in the necessity of reaching some solution or conclusion to a problem. The prejudiced child tends to "look too hard" in searching for a solution. Consequently, he becomes enmeshed in an attempt to win the approval of the examiner by finding any solution even to the extent of altering the problem and overstepping the conventions of logic.

However, tied to the prejudiced child's intolerance of ambiguity there appears in his cognitive functioning a *need for control* or mastery over the situation. In attempting to cope with a complex, uncertain, constantly changing environment, the prejudiced child tends to alter his world in keeping with a basic need to maintain stability, definiteness, and certainty. Hence, a need for control would seem to account for the positive, concretizing, dichotomizing, dogmatic, and simplifying approaches that typify his thought processes.

To what may we attribute such primary mechanisms of cognitive functioning as intolerance of ambiguity and a need for control? What role do these mechanisms play in the total personality organization of the prejudiced child? What are their origins? Below a suggested interpretation of the functional importance of these two central concepts is offered.

We have seen in previous sections that the prejudiced child tends toward a rigid cognitive structure. He tends to view his world in terms of dichotomies, categories, and unequivocalities. He is intolerant of ambiguity. He tends to think in concrete terms, "captured" by the stimulus field to such an extent that he is blind to existing field relationships. He tends to be maladaptive to new evidence and new circumstances requiring changes in behavior. He tends to overgeneralize and to dogmatically hold to his views in spite of the presence of contrary evidence. He tends to ignore discrepancies resulting from his cognitional distortions or to restructure them so as to produce a superficial degree of consistency. He attempts to be omniscient by avoiding the admission of a lack of knowledge.

Implicit in the analysis that follows is the assumption that cognitional functions are part-processes of the larger personality organization. Furthermore, these functions, it is implied, represent the fundamental motivational and emotional structures and needs of the individual in the cognitional sphere. Thus, Frenkel-Brunswick (14) refers to the cognitive and per-

ceptual factors that are part of "the more general factors in personality organization" (p. 461).

In order fully to comprehend our findings in the cognitive domain, it would be necessary to have insights into the nature of the child's adjustment within his family, his emotional needs and satisfaction, his fears and his pleasures. Such insights are provided by a study of the familial structure and family relationships in the homes of prejudiced and unprejudiced children carried out as part of the California Study. We shall attempt to relate our findings to those obtained through this latter investigation. Inasmuch as part of the basic findings of the California Study regarding early childhood experiences are founded upon inferences drawn directly from Ss' statements about their own lives, it would seem a legitimate procedure to assume (for the most part) the validity of these statements as indicative of the actual state of affairs. Frenkel-Brunswick emphasizes the fact that an S's statements may reflect the adult's preferred image of himself as a child. Further evidence from her study of late-childhood and adolescent prejudice, however, leads her to conclude that: "the descriptions which our subjects give of themselves [as children] show at least a certain degree of correspondence with the actual facts" (p. 435).

Typically, those high in ethnic prejudice came from families whose primary values were concerned with social status, standing in the community, being "normal," "conventional," "average." The family was hierarchically structured with the father as a stern and distant disciplinarian, and with family roles rigidly assigned: father as the provider, mother the homemaker; father the disciplinarian, mother the sacrificing and moralistic lady. Both parents were "moral-models": they established a pattern of morality for the family to follow and discipline was meted out primarily for violations of moralistic rules. Discipline was typically severe, ego-threatening, and overwhelming (as in threats of castration) and was imposed from without (as contrasted with an internalized super-ego in the subjects low in prejudice).

The home of the composite low-prejudiced person was characterized by policies of equality (as opposed to dominance and submission in the high-prejudice person), dependence for love or succorance and goals of child-training based upon the incorporation of principles of conduct rather than the imposition of arbitrary rules intended to "conventionalize" the child.

High-prejudiced persons tended to conform to the family-imposed standards while low-prejudiced persons were more critical of their parents in open opposition and showed a high degree of love-oriented identification.

Frenkel-Brunswik interprets the effects of these characteristic family patterns on the development of prejudice as follows:

"The lack of an internalized and individualized approach to the child, on the part of the parents, as well as a tendency to transmit mainly a set of conventional rules and customs, may be considered as interfering with the development of a clear-cut personal identity in the growing child. Instead, we find surface conformity without integration, expressing itself in a stereotyped approach devoid of genuine affect in almost all areas of life. . . . Even in the purely cognitive domain, ready-made clichés tend to take the place of spontaneous reactions" (p. 385).

By the time most children are seven years old they have already learned the game of social relationships in the context of the family unit. The more prejudiced seven-year-old has been forced to accept a set of comparatively rigid "laws" for getting along with his parents. On the whole, he does not comprehend the "reasons" that these laws exist but he does know that to violate them means punishment, often of a severe and ego-threatening nature. Hence, the more prejudiced child tends to submit to the rules and to "play the game" according to them. Failing to comprehend the ultimate purposes of the stringent restrictions placed upon him, he assumes that some *evil external* force makes some children "bad" and others "good." As Frenkel-Brunswik puts the matter, when the child's behavior violates the conventions of behavior ordained by the parent these conventions are rendered "ego-alien": they are not really part of his "normal" or "integrated" self but may be attributed to outside forces such as bad luck, or simply "forgotten."

Ego-alienation of other aspects of the child's personality structure becomes the rule: the child does not develop identifications within the family that are based upon affection. Rather, his relationships with parents and siblings may more properly be described as "associations." It is probable that the key to the child's ego-alienation is the unassimilable threat of punishment posed to the ego based upon what to the child is probably an incomprehensible and unpredictable "Damocles sword." Yet, he is still dependent upon his parents. Approval is given for adherence to the parental proscriptions. As a consequence of being both dependent upon and fearful of his parents, the prejudiced child tends to repress the hostility he feels against them. Such feelings would be wholly unacceptable in a situation demanding conformity. Ambivalence toward the parents becomes an anathema to the prejudiced child.

As in other cases of severe repression of un-

acceptable ideas or behavior, the child overcompensates with positive feelings for his parents: they are not only good parents, they are ideal. The California Study points up the exaggerated idealization and unrealistic evaluation of their parents by the more prejudiced group. Similarly, through a reaction formation, those violating the code of behavior dictated by the parents are seen as violators of societal codes: there is a violent rejection of nonconforming, flexible, or unconventional behavior. The seven-year-old prejudiced child rejects those who do not fit the patterns laid down as acceptable by his parents. In his terms the "value-violators" are those who are in some way different from himself and his ego-ideal.²² Typically, his is in an all-or-none response: the outgroup is rejected *en masse* and the ingroup is unquestionably accepted as a whole.

The cognitional correlate of the rejection of ambivalence, according to Frenkel-Brunswik, is intolerance of ambiguity. Associated closely with intolerance of ambiguity is a high degree of rigidity featured mainly by dogmatism, fixed set, and close adherence to convention.

We may now turn to the main task of the present section: attempting to define the functions that appear to be served by intolerance of ambiguity and a need for control in the ethnically prejudiced seven-year-old.

As noted above, the cognitive counterpart of emotional ambivalence is intolerance for ambiguity. The child avoids those aspects of his environment in ambiguous situations that pose a threat to him. Ambiguity threatens to bring to the surface of consciousness the basic anxiety due to the presence of ambivalence in the prejudiced child's personality organization. We have quoted Frenkel-Brunswik's view that ambivalence leads to dichotomization and displaced projected aggression.

We suggest that the prejudiced child's fear of ambiguity leads to a need for control: that is, he strives to impose structure upon ambiguity, decisiveness upon uncertainty. He attempts to control his environment so that the uncertain, indecisive, and unknown are eliminated. Particularly, he attempts to remove from experience equivocal aspects of his environment. Thus, he removes the enigma of diversity within ethnic outgroups. There cannot be some dirty Negroes, since the existence of an unknown group of potentially "different" Negroes casts doubt on the status of all Negroes. By asserting the universal rejection of outgroup members he can be sure that each Negro is "pegged." The need for controlling the

²² Frenkel-Brunswik (1, p. 430) has pointed out that the adult bigot's ego-ideal and image of the self are identical (i.e., he is everything he would want himself to be).

environment (as well as for controlling himself) helps the individual feel secure that his world is stable, permanent, predictable, rigid. The seven-year-old's rejections appear to be more "universal" than those of the adult bigot.²³ Hence, in the sphere of group attitudes, the child who is high in prejudice categorizes and dichotomizes his responses. The child achieves status through association in the identical fashion that status is achieved in the family group—by adhering to the prescribed code of standard behavior.

In the area of concept formation, the prejudiced child tends to become stimulus-bound, concrete in approach and random in seeking possible clues to the concept. He tends not to be capable of generalizing in a manner appropriate to the reality demands of the situation. Some of the more prejudiced Ss jump to conclusions while others offer unrelated "guesses" in the hope of hitting the right formula. The greater frequency of guesses and from other evidences (such as the almost complete lack of Don't Know responses on the DRT), suggests that the prejudiced child has a need for *pseudo-omniscience*, closely related to the need for control. If all is known, there is no fear of the unknown to be encountered and, similarly, complete knowledge obliterates any ambiguities that may in fact exist.

When inescapable cognitive ambiguity is encountered, the prejudiced child tends to render it harmless by removing the factors in the situation that are equivocal. In the Deductive Reasoning Test, he tended to avoid dealing with the relatively complex, seemingly contradictory ambiguous problems by segmenting them into simple, unambiguous "formal logic" problems, each with an autonomous solution. His responses, as a consequence, were more concrete than those of the unprejudiced child. Concretization and furation served to simplify complex phenomena for the child.²⁴ These phenomena are converted into understandable, manageable, discrete events.

In attempting to solve the conceptual prob-

lems of the problem-box test some of the prejudiced children tended to employ a "formula" to solve the problems. They developed set-superstitions and followed instructions in a concrete manner. The situation was controlled on the first critical problem by advancing a variety of inappropriate suggestions toward the solution. By the time the third critical problem was presented, the prejudiced child tended to become disturbed and sought in various ways to end the session. Failure resulted, in a large proportion of cases, not from the perseverative set that had contributed to failure on the first critical problem, but from the effects of low frustration-tolerance. The situation was no longer manageable but threatened to arouse feelings of anxiety. Hence, the child substituted "inadequate," random hypotheses, surface cues, or signs from E in an attempt to achieve an "accidental" solution.

It is because ambiguity, uncertainty, or indefiniteness threatens to arouse feelings of anxiety and hostility that lie below the surface of consciousness, that the child's thought processes become rigid. The rejection of hostility against the parents channelizes the affect associated with this hostility into external areas of life. To ensure that the feelings of hostility remain unavailable to consciousness, all areas of life in which ambivalences (or ambiguities) are manifested are rejected, converted, simplified, and hence controlled. Control over ambiguity results in control of unacceptable fear-arousing ambivalences.

Intolerance of ambiguity may be seen as a self-protective device whose function is to control ego-threatening unconscious feelings of hostility.

The child controls and manipulates his familial relationship successfully but outwardly conforming and submitting to the dominant parents. He generalizes his own need to conform and condemns others for failing to submit to the same pattern of rules of conduct to which he adheres. Those who fail to conform are universally rejected. Where the parents tend to emphasize the violations, the child will be quick to agree (i.e., "Don't play with colored children, they are bad" is generalized to "colored means bad," hence to the universal condemnation of Negroes).²⁵ Similar dichotomous moralisms feature the prejudiced child's views of people, objects, and situations. He improperly generalizes, rejects unknown or equivocal stimuli (as well as being disturbed by ambiguous life situations), and demands a clear, unhampered, stereotypical,

²³ The latter makes allowances for experience to modify his extreme views by asserting the existence of "exceptions" in the rejected ethnic groups.

²⁴ In similar fashion these mechanisms are employed by the adult bigot in accounting for "diverse" attributes of minority group members. Thus, for example, the prejudiced adult will speak of "the Negro" or "the Jew" in their generic senses and will bifurcate his views on Jews thus: "Jews are clannish, they always stick together" with "Jews are always trying to worm their way into non-Jewish groups."

²⁵ One parent reported that her child avows a hatred for "colored people." Upon interviewing the child E learned that the child's mother had warned him against "colored" children. He asserted to E, "I don't like colored people. I just like people who are black and white."

clichéd view of himself, his parents, and his world.

Obviously, these speculative interpretations require empirical testing. Among other possible fruitful research areas in this field would be studies relating parental methods in child-rearing to the development of social attitudes and to processes of thought. The results of the

present study seem to indicate that consistent patterns of coping with both the social and the natural world have their origins in childhood experiences. Investigations of the factors in the child's social experience tending to establish these patterns would be of great value to students of learning, personality, and intergroup relations.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ALLPORT, G. W. Prejudice; a problem in psychological and social causation. *J. soc. Issues*, 1950, Suppl. Ser. No. 4.
3. ALLPORT, G. W. *The nature of prejudice*. Cambridge, Mass.: Addison-Wesley, 1954.
4. ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, 22, 9-39.
5. BLAKE, R. R., & RAMSEY, G. V. *Perception: an approach to personality*. New York: Ronald Press, 1951.
6. BOLLES, M. M. The basis of pertinence. *Arch. Psychol.*, N.Y., 1937, No. 212.
7. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, 42, 33-44.
8. BRUNER, J. S., & KRECH, D. W. (Ed.) *Perception and personality*. Durham, N.C.: Duke Univ. Press, 1949.
9. BURT, C. *Mental and scholastic tests*. London: King, 1921.
10. COWEN, E. L., & THOMPSON, G. G. Problem solving rigidity and personality structure. *J. abnorm. soc. Psychol.*, 1951, 46, 165-175.
11. CRISWELL, J. H. A sociometric study of race cleavage in the classroom. *Arch. Psychol.*, N.Y., 1939, No. 235.
12. FRENKEL-BRUNSWIK, ELSE. A study of prejudice in children. *Human Relat.*, 1948, 1, 295-306.
13. FRENKEL-BRUNSWIK, ELSE. Dynamic and cognitive categorization of qualitative material: I. General problems and the Thematic Apperception Test; II. Application of interviews with the ethnically prejudiced. *J. Psychol.*, 1948, 25, 253-277.
14. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual variable. *J. Pers.*, 1949, 18, 108-143.
15. FRENKEL-BRUNSWIK, ELSE. Patterns of social and cognitive outlook in children and parents. *Amer. J. Orthopsychiat.*, 1951, 21, 543-558.
16. GESELL, A., CASTNER, B. M., THOMPSON, H., & AMATRUDA, C. S., *Biographies of child development: the mental growth careers of eighty-four infants and children*. New York: Hoeber, 1939.
17. GOODMAN, M. E. *Race awareness in young children*. Cambridge, Mass.: Addison-Wesley, 1952.
18. GRAHAM, J. L. Learning to generalize. *Psychol. Monogr.*, 1938, 50, No. 5 (Whole No. 225), 84-115.
19. HANFMANN, EUGENIA, & KASANIN, J. A method for the study of concept formation. *J. Psychol.*, 1937, 3, 521-540.
20. HAYAKAWA, S. I. *Language in action*. New York: Harcourt, Brace, 1947.
21. HOCHBERG, J. E., & GLEITMAN, H. Towards a reformation of the perception-motivation dichotomy. *J. Pers.*, 1949, 18, 180-191.
22. HOROWITZ, E. L. The development of attitude toward the negro. *Arch. Psychol.*, N.Y., No. 194, 1936.
23. HOROWITZ, E. L., & HOROWITZ, R. E. Development of social attitudes in children. *Sociometry*, 1938, 1, 301-338.
24. ISAACS, S. *Intellectual growth in young children*. London: Routledge and Kegan Paul, 1930.
25. JAENSCH, E. R. *Der gegentypus*. Leipzig: Barth, 1938.
26. LASKER, B. *Race attitude in children*. New York: Holt, 1935.
27. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of intensity of a need to the amount of perceptual distortion. *J. Psychol.*, 1942, 13, 283-293.
28. LONG, L., & WELCH, L. Reasoning ability in young children. *J. Psychol.*, 1941, 12, 21-44.
29. MCGINNIES, E., & BOWLES, W. Personal values as determinants of perceptual fixation. *J. Pers.*, 1949, 18, 224-235.

30. MELTZER, H. The development of children's nationality preferences, concepts and attitudes. *J. Psychol.*, 1941, 11, 343-358.
31. MINARD, R. D. Race attitudes of Iowa children. *Univer. Iowa Stud.: Stud. Character*, 1931, 4, No. 2.
32. MORENO, J. L. *Who shall survive?* Washington, D.C.: Nervous & Mental Disease Pub. Co., 1934.
33. MORGAN, J. J. B., & MORTON, J. T. The distortion of syllogistic reasoning by personal convictions. *J. soc. Psychol.*, 1944, 20, 39-59.
34. MURPHY, G. Foreword to E. L. Hartley, *Problems in prejudice*. New York: King's Crown, 1946.
35. MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
36. NEWCOMB, T. M. *Personality and social change*. New York: Dryden, 1943.
37. OAKES, M. E. Children's explanations of natural phenomena. *Teach. Coll. Contr. Educ.*, 1946, No. 926.
38. O'CONNOR, PATRICIA. Ethnocentrism, "intolerance of ambiguity," and reasoning ability. *J. abnorm. soc. Psychol.*, 1952, 47, 526-530.
39. PIAGET, J. *Judgment and reasoning in the child*. New York: Harcourt, Brace, 1928.
40. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, 43, 142-154.
41. RADKE, M., TRAGER, H. G., & DAVIS, H. Social perceptions and attitudes of children. *Genet. Psychol. Monogr.*, 1949, 40, 327-447.
42. REVESZ, G. An experimental study in abstraction in monkeys. *J. comp. Psychol.*, 1925, 5, 293-343.
43. ROGERS, K. H. Perseveration. *J. ment. Sci.*, 1935, 81, 138-144.
44. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1943, 48, 259-278.
45. ROKEACH, M. A method for studying individual differences in "narrow-mindedness." *J. Pers.*, 1951, 20, 219-233.
46. ROKEACH, M. "Narrow-mindedness" and personality. *J. Pers.*, 1951, 20, 234-251.
47. ROKEACH, M. Prejudice, concreteness of thinking, and reification of thinking. *J. abnorm. soc. Psychol.*, 1951, 46, 83-91.
48. ROKEACH, M. The nature and meaning of dogmatism. *Psychol. Rev.*, 1954, 61, 194-204.
49. ROKEACH, M., MCGOVNEY, W. C., & DENNY, M. R. A distinction between dogmatic and rigid thinking. *J. abnorm. soc. Psychol.*, 1955, 51, 87-93.
50. SELLS, S. B. The atmosphere effect. *Arch. Psychol.*, N.Y., 1936, No. 200.
51. SELLS, S. B., & KOOB, H. F. A classroom demonstration of "Atmosphere effect in reasoning." *J. educ. Psychol.*, 1937, 28, 514-518.
52. SMITH, M. B. Personal values as determinants of a political attitude. *J. Psychol.*, 1949, 28, 477-486.
53. SPEARMAN, C. E. *Abilities of man: their nature and measurement*. New York: Macmillan, 1927.
54. THOMPSON, J. The ability of children of different grade levels to generalize on sorting tests. *J. Psychol.*, 1941, 11, 119-126.
55. TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, 55, 189-208.
56. VICKERY, W. E., & OPLER, M. E. A redefinition of prejudice for purposes of social science. *Hum. Relat.*, 1947-1948, 1, 419-428.
57. WEIGL, E. Zur Psychologie Sogenannter Abstraktionsprozesse: I. Untersuchungen über das "Ordnen." *Z. Psychol.*, 1927, 103, 1-45.
58. WERNER, H. *Comparative psychology of mental development* (rev.). Chicago: Follett, 1948.
59. WILKINS, M. C. The effect of changed material on ability to do formal syllogistic reasoning. *Arch. Psychol.*, N.Y., 1928, No. 102. Pp. 5-83.
60. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
61. WOODWORTH, R. S. & SELLS, S. B. An atmosphere effect in formal syllogistic reasoning. *J. exp. Psychol.*, 1935, 18, 451-460.
62. ZELIGS, ROSE. Racial attitudes of 200 sixth-grade children. *Sociol. soc. Res.*, 1933-1934, 18, 26-36.
63. ZELIGS, ROSE. Tracing racial attitudes through adolescence. *Sociol. soc. Res.*, 23, 1938, 45-54.
64. ZELIGS, ROSE. Children's intergroup attitudes. *J. gen. Psychol.*, 1947, 72, 101-110.
65. ZELIGS, ROSE, & HENDRICKSON, D. Factors regarded by children as the basis of their social attitudes. *Social soc. Res.*, 1934-1935, 19, 225-233.

(Accepted for publication February 20, 1957)

APPENDIX A

ETHNIC ATTITUDE TEST

1. There are city people and there are people who live in the country on farms. Some city people say that the farmers are pretty dumb and they laugh at farmers. What do you think about that? How do you feel about farmers?
2. Some children learn slowly and some children learn fast. The fast children say that the slow ones are stupid or dumb. They don't like to make friends with slow children. What about children who learn slowly?
3. Some people think that Mexicans are not very nice. They say that Mexicans are dirty; that they are dumb and that they are not as good as we are. What do you think of that? How do you feel about Mexicans?
4. Some people are Jewish and some people are Catholics. Some Jewish people say that the Catholic people are not very nice. They say that Jewish is a better religion than Catholic. They don't like to make friends with Catholic people. What about that? How do you feel about Catholic people?
5. Some people are poor and some people are rich. Some people say that poor people would not be poor if they weren't so lazy and dumb. What do you think of that? What about poor people? Some people say that rich people made their money by not being honest, that they cheated other people and became rich. What about that? How did rich people get their money?
6. There are colored people and there are white people. Some white people say that colored people are bad, that they steal and they are dirty and lazy and smell bad. What about colored people? How do you feel about colored people?
7. Some people say that Jewish people are not very nice. They say that Jewish people are too rich and that they are not like other people. They say that Jewish people are cheaters and they do not want them for friends. What about Jewish people? How do you feel about Jewish people?
8. Some people say that Americans are better than any other people in the whole world. They say that American people do things better and make nicer things than other people. How about Americans and the other people? Are Americans better than other people?
9. Some people say that white people are not nice to have as friends. They say that white people are not as nice as colored people. They say that white people are mean and hurt people and fight with them. What about white people?

Control of Behavior in Chimpanzees and Pigeons by Time Out from Positive Reinforcement^{1,2}CHARLES B. FERSTER³*Yerkes Laboratories*

THIS series of experiments is an analysis of some properties of behavioral control by stimuli which are aversive because positive reinforcement is discontinued. The basic paradigm used in these investigations for studying this kind of control uses a "time out" from positive reinforcement as the aversive event. To study the effect of the time out, a simple response, such as a pigeon's peck at a circular disk, or a chimpanzee's press of a toggle switch, was maintained when occasional pecks or presses were followed with food. Reinforcements became available on a variable schedule, which produced sustained responding at substantial rates. The resulting performance then served as a base line on which to assess the effects of the introduction of the aversive stimuli. To generate the aversive event, reinforcement was withheld in the presence of one stimulus but alternately continued in the presence of a second stimulus. When the rate of responding fell to zero in the presence of the stimulus where reinforcement was discontinued, that stimulus, called the time-out stimulus, could then be used as

an aversive event and its effect on the base-line performance measured.

The "time out" is useful for interrupting an experiment, since all that is required is that a stimulus be changed. Hence, the time out has been used:

1. To begin and end experimental sessions (6). A time out at the beginning of the experiment enables effects of handling the animal to disappear before the experiment is started; a time out at the end of the experimental session terminates the daily session without the presence of the experimenter.

2. In drug experiments, by interrupting the experiment with a time out and continuing the experiment some time later when the drug has had its maximum effect (3).

3. As a way of analyzing and controlling the temporal properties of the animal's past behavior on a current performance (7). For example, the discriminative effects of the reinforcement can be minimized if it is followed by a time out, so that, when the experiment is resumed, a period of time has elapsed during which the reinforcement would have been effective.

A time out from positive reinforcement has been used to punish responses in the development of a discrimination, or differentiation. The aversive properties of a time out were also evident in experiments where a time out scheduled during a fixed-ratio performance disrupted the performance severely, because the high fixed-ratio rate of responding resulted in an inevitable correlation of the time out with a response (7).

Finally, Morse and Herrnstein, using pigeons, demonstrated that behavior could be maintained because it postponed a time out (13).

The primary aversive stimuli have relatively fixed effects. In contrast, the effect of the time out is determined by the schedule of reinforcement it interrupts. At extreme values of some schedules of reinforcement, the time out may become positively reinforcing instead of aversive.

¹ This study is part of a project supported by the United States Atomic Energy Commission under Contract AT-(40-1)-1553.

² The part of Experiment IV using pigeons was carried out in the Department of Pharmacology of the Harvard Medical School. Thanks are due to Otto Kraymer and P. B. Dews for making the facilities available.

³ Now at the Institute for Psychiatric Research, Indiana University Medical Center.

Morse and Herrnstein studied this property of the time out in the extreme case by experiments in which pigeons were reinforced on (a) large values of fixed-interval schedules, or (b) schedules in which low rates were differentially reinforced; that is, responses were reinforced only when they were spaced at least five minutes from a preceding response (9). They found that the stimuli associated with these schedules were aversive, since the pigeon had a strong disposition to peck a second key which interrupted the experiment for a long interval of time (e.g., three minutes).

Most of the punishments in human affairs involve the withdrawal of positive reinforcements (time out) rather than primary noxious events, or aversive stimuli. Jails, discharge from employment, and ostracization are a few of the more extreme forms. Even when corporal punishment is used, as for example in parental control of children, its effect is confounded with time out from positive reinforcement. A parent who spansks a child is less likely to provide the normal positive reinforcement.

Because the function of the time out depends (at least at extreme values) on the schedule of reinforcement from which it is derived, a single schedule of reinforcement and body weight was used in all of the experiments reported here. (An exception was made in the part of Exp. III using pigeons.) A variable-interval schedule with a mean of three minutes (VI₃) was chosen as a moderate value. The average interval between reinforcements was three minutes, and the actual intervals were randomized between zero (successive responses) and six minutes. Such a variable-interval schedule will normally produce a moderate and near-constant rate of responding. Furthermore, the experimental session is long

enough to impose repeated stimulus conditions, and also to minimize the relative importance of behaviors associated with too frequent eating. At the same time, the reinforcements occur frequently enough to prevent the schedule of reinforcement from becoming aversive.

SUBJECTS

The subjects of the experiments were two adult chimpanzees, Yerkes Colony No. 67 (Jed) and No. 97 (Verb), 10 and 15 years old, respectively. The animals had a prior history in various manual discrimination experiments. In Exp. III, the subjects were two White Carneaux male pigeons, approximately two years old. They had a history of variable-interval reinforcement, during which their weights were varied.

APPARATUS AND PROCEDURE

The principles used in the design of the apparatus were described elsewhere (6).

The chimpanzees' living quarters consisted of a room (12 by 12 feet) adjoining a fenced enclosure (15 by 15 feet). The apparatus was permanently mounted in the indoor portion (the experimental chamber). While one subject took part in the experiment, the second subject was locked out in the fenced enclosure. A scale permanently mounted from the ceiling of the experimental cage was used to weigh the animals before each experimental session. The animals were trained to hang quietly from the scale by successive approximation of the final behavior (18). Then, hanging from the scale was reinforced only when an appropriate light was on, so that the animals could be weighed at any time simply by turning on the light. The behavior recorded was the operation of a telephone-type key, 30 inches from the floor. The food magazine delivered alternate 50-calorie portions of hard whole wheat cracker and Purina monkey chow into a food tray 12 inches to the left of the key. Control and recording equipment were kept in an adjoining room.

Number 97 entered the experimental chamber at 9:30 A.M. and remained until 4:00 P.M.; No. 67 entered at 4:30 P.M. and remained until 8:00 A.M. The animals were kept in the same

room between 8:00 and 9:00 A.M., when both compartments were cleaned. The beginning and end of the experimental sessions were marked by the presentation and withdrawal of an overhead light. The daily session was terminated when the animal received 55 to 60 reinforcements, or when eight hours elapsed, whichever came first. The absence of the light controlled a near-zero rate of key pressing, because of differential reinforcement before the start of the experiment.

The animals were maintained at approximately 80% of their free-feeding body weight. They were fed a supplement of yeast tablets, two oranges, and four ounces of evaporated skim milk after the experimental session, when they went to the outside part of the living quarters. Any day when an animal's weight exceeded its designated experimental weight, the experimental session was postponed and the animal fed only supplementary rations. Extra food was given after the experimental session whenever the subject's weight fell below the designated experimental weight for several days.

The time out was programmed by shutting off the lights in the experimental room and disconnecting the reinforcing circuits. Any response occurring during the time out extended its full duration. The recorder paper feed did not run during the time out. Any responses occurring would be noted by a vertical rise in the recording pen. The rate of responding remained essentially zero during the time out. In all of the experiments the time out lasted 10 minutes except in the occasional instances where a response during the time out prolonged it. The intervals in the variable-interval schedule averaged three minutes and ranged from several seconds to six minutes in an arithmetic progression, except for a slightly lessened density of short intervals. The variable-interval schedule used with the pigeons in Exp. III averaged six minutes, varying from several seconds to 12 minutes.

The apparatus used with the pigeons in Exp. III is described elsewhere (6). No reinforcement could occur when a response produced a time out.

The actual order in which the experiments were carried out was I, II, IV, V, and III.

EXPERIMENT I. THE PRE-TIME-OUT STIMULUS

The suppression of behavior by a stimulus preceding an aversive event (pre-shock, pre-time-out, or pre-aversive stimulus) is an important case in a survey of the properties of an aversive stimulus. This is the experimental paradigm first

demonstrated by Estes and Skinner (5) and later extensively used by Brady and Hunt (2). Herrnstein, extending the pre-aversive procedure to the time out, found that the pre-time-out stimulus suppressed behavior only when the reinforcement was given very frequently—e.g., a variable-interval schedule with a mean of 30 seconds. At higher mean values of the variable-interval schedule, however, the pre-time-out stimulus controlled a higher rate of responding than the prevailing interval rate. Herrnstein interpreted this increased rate as positive reinforcement by the time out because of the aversive properties of the schedule of positive reinforcement (8).

This experiment investigates the pre-time-out procedure in the chimpanzee at a moderate value of the variable-interval schedule, chosen to emphasize the aversive properties of the time out rather than the schedule of reinforcement.

Procedure

The subjects were reinforced on a variable-interval schedule for 65 hours following one session of continuous reinforcement. The 30-second, pre-time-out stimulus was a red lamp just above the key. The red light appeared at various intervals, averaging every 10 minutes. Reinforcements occurred on the three-minute, variable-interval schedule, independently of the red light. The red light was not followed by the time out for the first two sessions of the experiment. On the third session, the first response 30 seconds after the appearance of the red light produced a time out that interrupted the experiment for 10 minutes. During the remainder of the experiment, the effect of the red light on the rate of responding was measured variously with and without the time out following the red light. To check for artifacts, control sessions were carried out in which the red light was disconnected, but all of the recording circuits were still in effect.

Results

Figure 1 contains summary graphs of the whole experiment for both animals showing the ratio of the rate of respond-

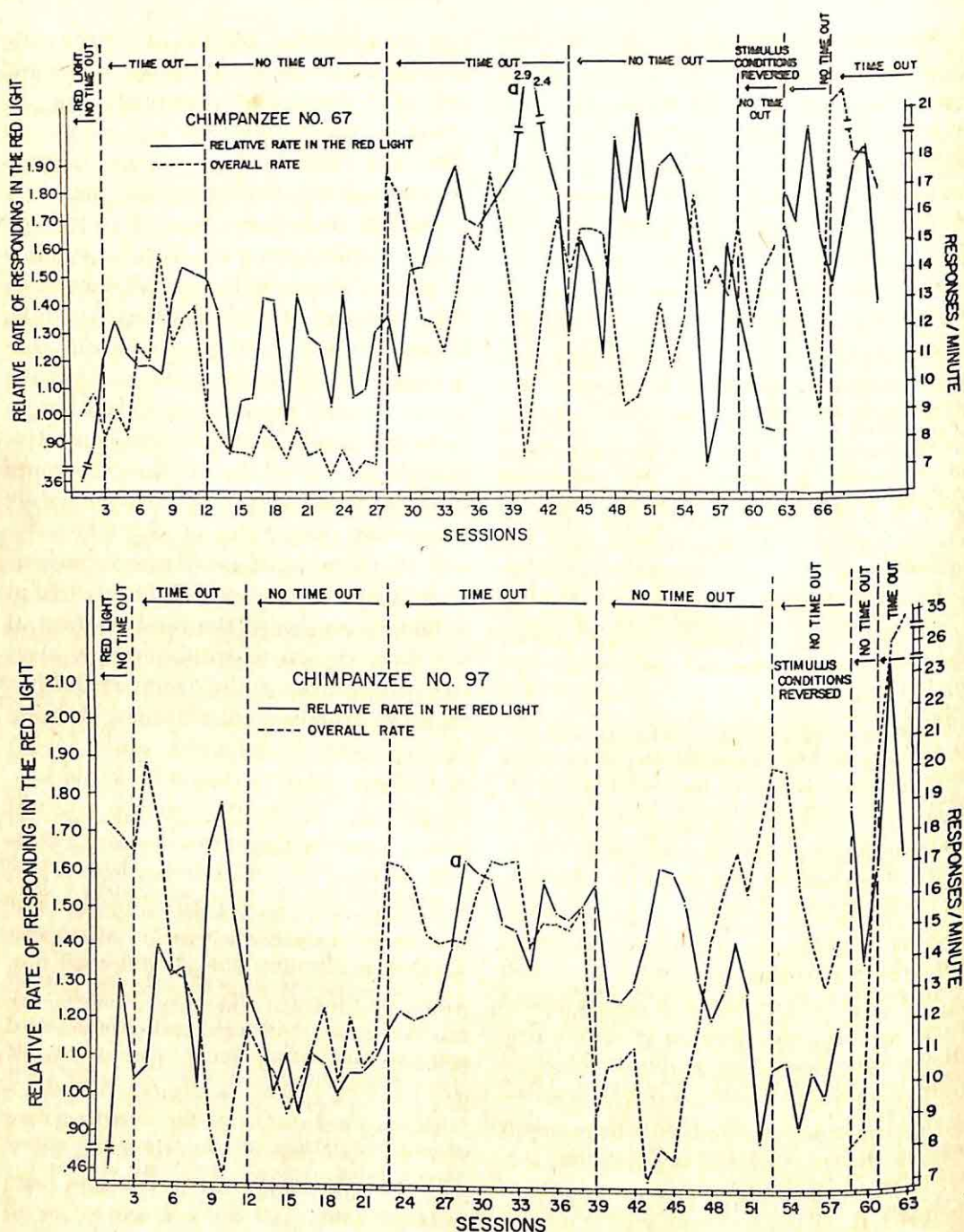


FIG. 1. Chimpanzees No. 67 and 97. The relative rate of responding during the pre-time-out stimulus and over-all rate of responding. Each point is the average of an entire daily session.

ing during the pre-time-out stimulus to the over-all rate. The time out increases the rate of responding in the presence of the pre-time-out stimulus, although the

effect is not completely reversible. The highest relative pre-time-out rates of responding occurred when the over-all rate was low.

Rate of responding during the pre-time-out stimulus relative to the over-all rate of responding. Before the time out was introduced, the rate during the red light was low for both animals. After the time out was introduced, the rate of responding in the presence of the red light increased for both animals and remained considerably above the over-all rate. When the time out was discontinued in Session 12, the rate of responding during the red light fell for both animals. Number 97's rate quickly reached a value approximating the general level of responding. Number 67, however, showed many sessions in which the rate of responding during the red light increased to values as high as the previous procedure with the time out. When the time out again followed the red light, No. 67's pre-time-out rate increased markedly, reaching approximately 1.7 times the over-all rate; and on two sessions it reached 2.9 and 2.4 times the over-all rate. Number 97 also showed an increase over the pre-time-out rate of responding; but this increase was not so marked as that of No. 67. The rate during the red light reached approximately the same value as during the previous time-out procedure.

When the time out was discontinued for a second time, the rate of responding during the 30-second period of red light remained high. To control for possible artifacts, sessions were carried out in which the red light was disconnected, with the recording program remaining as before. Both animals showed rates of responding approximately equal to the over-all rate.

To investigate the continuing control of a high rate by the red light in the absence of the time out, the stimulus conditions were then reversed. The red light remained on for the bulk of the session,

and the pre-time-out stimulus became a 30-second period in which the red light was off. The rate of responding during the 30-second intervals in which the red light was off fell relative to the over-all rate during the rest of the session when the red light was on. The rate fell to the value of the over-all rate, but not below it. When the original stimulus conditions were restored (the time out still not following the red light), the rate of responding again increased, reaching the level present before the reversal of the stimulus conditions. For No. 97, the rate during the red light then reached the highest rate of responding yet observed.

Over-all rate of responding. The over-all rate of responding, shown in the second curves of Fig. 1, shows a relation both to the introduction and removal of the time out and the relative rate of responding during the red light. The over-all rate of responding fell when the time out was discontinued (Sessions 12-27) and rose markedly when it was reintroduced (Session 28). The second removal of the time out had no obvious effect, but the reintroduction of the time out on Session 66 was accompanied by another marked rate increase. Number 97, in Fig. 1, showed a similar effect of the time out. The over-all rate increased on Session 23, when the time out was reintroduced, and fell on Session 39, when it was discontinued. The rate fell again on Session 59, with no time out, and increased on Session 61, with its reintroduction.

The highest relative pre-time-out rates of responding occurred when the over-all rate was low. For No. 67, on Sessions 33, 40, 50, and 65, peak values of the pre-time-out rate of responding occurred with the lowest values of the over-all rate of responding. Conversely, when the over-all rate reached its peak values, the relative rate during the red light tended

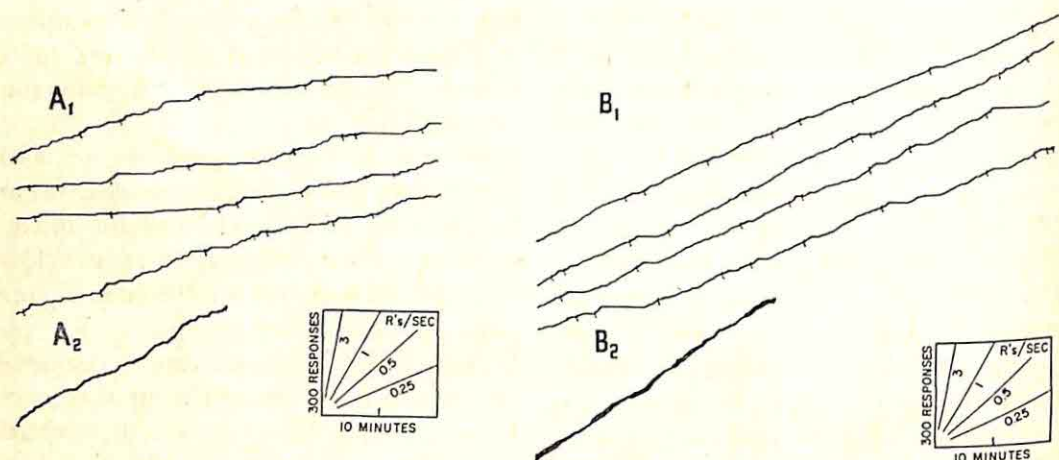


FIG. 2. Chimpanzees No. 67 and 97. Cumulative curves of an entire daily session showing the highest pre-time-out rates of responding relative to the over-all rate. Record B contains the performance during the pre-time-out stimulus only.

to fall. Note, for example, that the two curves cross frequently. Number 97 (Fig. 1) shows a similar performance. Peak values of the relative pre-time-out rate are associated with the lowest values of the over-all rate (Sessions 10, 44, and 60). However, there is no relation between high over-all rates and low values of the relative pre-time-out rate, as with No. 67.

Figure 2 contains the cumulative curves of an entire daily session, showing the performances for both animals, when the highest relative pre-time-out rates of responding were observed (at the points marked *a* in Fig. 1). In each case, Record 1 contains the entire performance, and Record 2 contains only the behavior during the pre-time-out stimulus. The diagonal marks indicate the start of the 30-second period during which the red light was on; reinforcements and the position of the time out are not marked on the curves. The grid in the lower right part of the figure gives the coordinates, the scale of the record, and several reference slopes in responses per second.

In Records A-1 and A-2, the 40th session for No. 67, the pre-time-out rate

was 2.9 times as high as the over-all rate. The over-all rate of responding was low, with highest rates of the session occurring during the pre-time-out period. Occasional low rates appeared during the pre-time-out stimulus, as well as high rates elsewhere. The performance for No. 97 in Records B-1 and B-2 was similar, except for a higher over-all rate, showing more sustaining responding. The record contains the 29th session, where the pre-time-out rate was 1.6 times the over-all rate. Here, the rates of responding during the pre-time-out stimulus were frequently higher than any appearing elsewhere. Occasional instances occurred where the rate during the red light was lower than the prevailing rate.

When the pre-time-out rate was low relative to the over-all rate, the absolute values of the pre-time-out rates of responding were still substantial. In the main, the effect of the pre-time-out stimulus was to sustain responding rather than increase the local rate.

Discussion

The pre-time-out procedure produces an increase in rate of responding during

the pre-time-out stimulus, rather than a decrease, as with the preshock stimulus. In this experiment the increased rate during the pre-time-out stimulus cannot be interpreted as behavior reinforced by an aversive variable-interval schedule of reinforcement. Moderate schedule values were used; and, as Experiments III, IV, and V show, the interruption of the same schedule of reinforcement in the same animals is a highly aversive event.

Examination of the temporal relations of responses during the pre-time-out stimulus, the reinforcement, and the time out may suggest a basis for the increase in rate during the pre-time-out stimulus. The likelihood that a given response will produce a time out is high after a pause because the time out occurs at a fixed interval following the appearance of the red light. In the extreme case a response spaced 30 seconds from a previous response will always produce a time out. Conversely, only a small fraction of responses emitted at higher rates produces time outs. The result is a differential punishment of spaced responding because spaced responses are more likely to produce a time out. Also, the likelihood of a reinforcement following long pauses is low. Pauses of 30 seconds or more during the pre-time-out stimulus are never reinforced. This contingency is somewhat similar to schedules of reinforcement where the reinforcement is "set up" for a brief period of time (limited hold) rather than the normal procedure of the reinforcement being held indefinitely (7). The above contingencies produce large effects when programmed explicitly. Experiment V shows that the selective punishment of spaced responding is effective in controlling the rate of responding. Experiments with pigeons and rats have shown effective

control of the rate of responding when reinforcement is withheld until a response occurs at the designated value (16, 17, 19).

The fact that the pre-time-out rate of responding generally did not exceed the highest prevailing rate elsewhere during the session also suggests that a major effect of the time out is to eliminate long pauses.

Both animals confirm the effect of the time out in increasing the over-all rate of responding (see Exp. II). Therefore, the effect of the time out might not be completely under the control of the red light, or the time out may have effects other than the contingencies in the red light described above. Because the time out following the red light produces sustained responding at the highest prevailing rate during the pre-time-out stimulus, its effect is more marked at lower over-all rates of responding. This confirms Herrnstein's finding that the relative rate during the pre-time-out stimulus increased when the frequency of reinforcement was decreased. The over-all rate of responding declined as reinforcements occurred less frequently.

The failure of the pre-time-out rate of responding to return to the level of the base-line rate after the time out no longer occurred indicates a large order of magnitude of control. There was no indication that the pre-time-out rate was falling during the 23 sessions in which the time out no longer followed the 30-second, red-light period. The apparent irreversibility of the process might be due at least partially to the fact that the discontinuance of the time out did not provide any new contingency opposed in effect to the time out.

As in Experiments III, IV, and V, No. 67 showed a greater sensitivity to the effects of the time out than did No. 97.

EXPERIMENT II. THE EFFECT OF A PERIODIC TIME OUT ON THE OVER-ALL RATE OF RESPONDING ON VARIABLE-INTERVAL REINFORCEMENT

The effect of a time out in increasing the rate of responding on a variable-interval schedule was confirmed during the preliminary conditions of the avoidance experiment reported in Exp. IV. The avoidance procedure of Exp. IV produced a condition where the performance on the variable-interval schedule of food reinforcement was periodically interrupted by a time out, because no behavior had yet developed on the avoidance lever to postpone time outs. The periodic interruption of the experiment produced a marked increase in the over-all rate of responding. The avoidance experiment was then postponed so that the effect of the time out on the variable-interval rate might be examined in more detail.

Procedure

The performance on the variable-interval schedule was compared with and without time out. In the time-out procedure the subjects had access to the three-minute, variable-interval food reinforcement schedule for 60-second intervals, separated by three minutes when the time-out stimulus was present. The time-out procedure was compared with a control procedure of three-minute, variable-interval reinforcement without interruptions. In chimpanzee No. 67 the effects of time out and no time out were studied in alternate sessions for the first four sessions. The fifth and sixth sessions began without time out for the first hour, followed by the time-out procedure for the remainder of the session. In chimpanzee No. 97 the time out was studied in alternate sessions for the first two sessions. For the remaining four sessions of the experiment the procedure changed after the first hour of the session. In some cases the session began with time outs during the first hour, and in others with no time out. The previous history of the subjects is recorded in Exp. I.

Results

The rate of responding with the time-

out procedure was always higher than without the time out. The ratio of rates of responding under time-out and no-time-out conditions ranged between 5.1 and 2.3 for No. 67 and 3.7 and 1.6 for No. 97.

Records A and B of Fig. 3, which contain the entire fourth and fifth sessions for No. 67, show a more detailed account of the effect of the time out. The recorder paper feed stops during the time out. Reinforcements are marked by the diagonal lines on the record. In Record A the session began with three-minute, variable-interval reinforcement, with time outs every minute. After one hour of variable-interval reinforcement (at the arrow), time outs were discontinued for the remainder of the session. The rate of responding fell within 10 minutes and showed a continuing decline through the remainder of the session. On the following day, shown in its entirety in Record B, the session began with variable-interval reinforcement, with no time outs. After an hour (at the arrow), the time-out procedure was reinstated for the remainder of the session. The rate of responding increased as soon as the time out was added. Pauses up to 30 seconds occur after many time outs. These pauses following time outs also occurred in Exp. III, and have been previously reported by Herrnstein in other experiments using variable-interval reinforcement and time outs with a pigeon (8).

Number 97, the animal with the higher over-all rate, showed the same effect of the time out. The first discontinuation of the time outs was immediately followed by a fall in rate. On the following day, when the time out was reintroduced after an hour, the rate of responding increased, interrupting the progressive decline in rate that is a usual characteristic of variable-interval per-

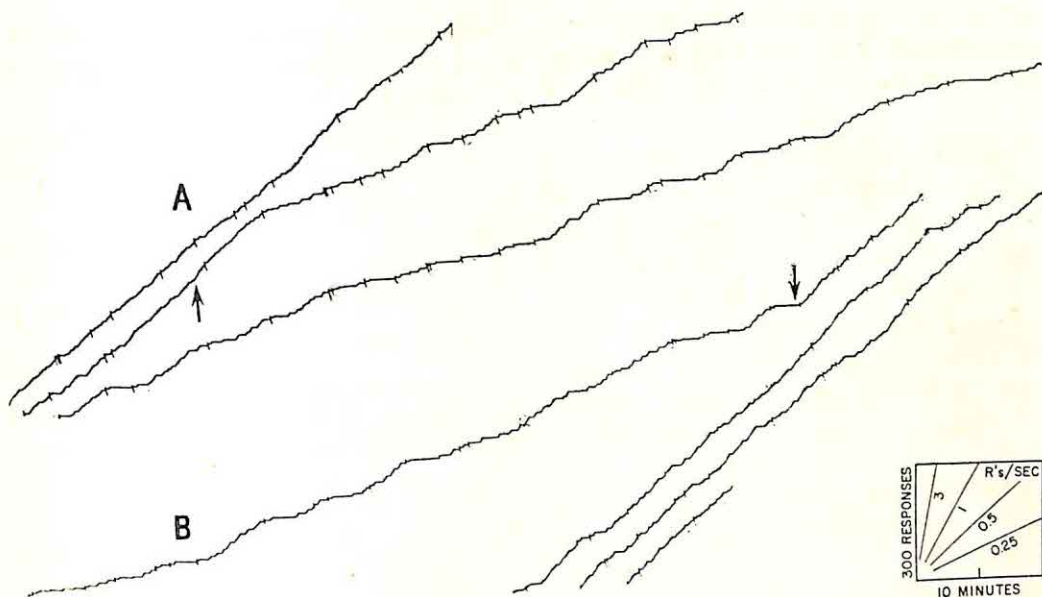


FIG. 3. Chimpanzee No. 67. Record A: The session is interrupted every minute. At the arrow the session is no longer interrupted. Record B: The following session. Beginning at the arrow, the session is interrupted every minute.

formances. The pauses following time outs were more marked for this animal than for No. 67.

The periodic time out has similar effects when the entire session is interrupted. Figure 4 contains two successive daily sessions for No. 97, comparing the performances with and without time out. Record A, without time out, shows a low over-all rate of responding with marked negative curvature during the session,

reaching near zero by the end of the session. On the following day (Record B) the periodic interruption of the experiment increased the over-all rate by a factor of more than two. The increased over-all rate was caused by sustained responding rather than higher local rates of responding. Segments at *a*, *b*, *c*, and *d*, in Record A, for example, were emitted at the prevailing rate of Record B. Brief pauses follow most of the time

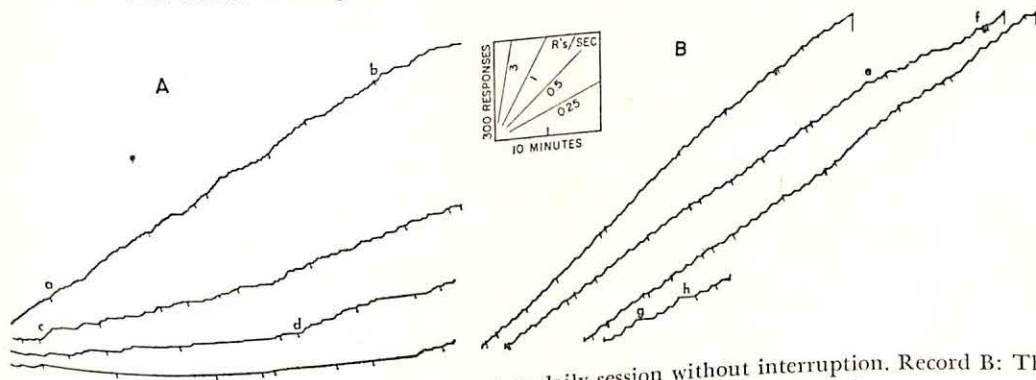


FIG. 4. Chimpanzee No. 97. Record A: A complete daily session without interruption. Record B: The session following Record A, where the session is interrupted every minute.

outs, as the wavy character of the record indicates. They are especially marked in the region *e* to *f* and at *g* and *h*.

Discussion

The increased rate of responding when the variable-interval reinforcement procedure was periodically interrupted confirms the effect of the time out on the rate of responding during the pre-time-out stimulus in Exp. I. The lesser amount of food per unit of time under the time-out conditions confounds the result somewhat; but the rate of food ingestion is probably not a large enough factor to produce the order of magnitude of rate changes reported here. Under the highest rate of food delivery, reinforcements are still spaced on an average of every three minutes. Contingencies are present in this experiment (similar to those already described in Exp. I) which could be responsible for the increased rate of responding when the session is interrupted periodically. The likelihood of a time out occurring is higher after longer pauses than shorter pauses. Hence, the class of reinforced responses will be partially restricted in the direction of responses spaced closely, because a higher proportion of distantly spaced responses will produce time outs (hence no reinforcement) than closely spaced responses. This lower frequency of reinforcement is tantamount to the reinforcement of higher rates. The contingency here is also similar to that of reinforcement schedules where the reinforcement is "set up" for a brief period rather than being available until the next response (limited-hold reinforcement). The reinforcement on a limited-hold basis reduces the likelihood of reinforcement after long pauses, just as in the time-out procedures of Exp. I and II, and increases the rate of responding.

EXPERIMENT III. THE SUPPRESSION OF BEHAVIOR BY A STIMULUS PRECEDING A TIME OUT FROM A SCHEDULE OF POSITIVE REINFORCEMENT

In this experiment the aversive properties of the time out were emphasized by terminating the pre-time-out stimulus, without a time out, if the rate of responding were low enough during its latter part. Pauses initiated during the pre-time-out stimulus reduce the likelihood of a time out and make a reinforcement more probable. Responses emitted at higher rates during the pre-time-out stimulus, however, are more likely to be punished. The contingencies are just opposite to those of Exp. I, where the first response 30 seconds after the appearance of the pre-time-out stimulus is followed by a time out. There, a pause initiated during the pre-time-out stimulus made punishment more likely. As a control procedure, the frequency of reinforcements on the variable-interval schedule was decreased to every six minutes, on the average, a value at which Herrnstein's experiment showed increased rates of responding during the pre-time-out stimulus. The experiment was first carried out with pigeons and then confirmed with the chimpanzees of Exp. I and II.

Pigeons

Procedure

The subjects of the experiment were two adult pigeons, maintained at approximately 80% of their free-feeding weight. Pecks at a translucent key were reinforced on a six-minute, arithmetic variable-interval schedule of reinforcement. The longest interval was 12 minutes and the shortest, several seconds. In order to maintain the low frequency of reinforcement without an excessively long programming tape, a few of the shorter intervals were omitted. The experiment began with a procedure similar to that of Exp. I. A blue light appeared behind the key every 11 minutes, and the first response 30 seconds after the appearance of the blue light (pre-time-out

stimulus) terminated the experiment for 10 minutes. The procedure was then changed so that the blue light went off without a time out if the rate of responding were low enough during the last part of the pre-time-out stimulus. This was done by arranging a short period of time in the last part of the pre-time-out stimulus (two to five seconds) when a peck would produce a time out. If no response occurred, the stimulus terminated without a time out. The level of deprivation, the length of the pre-time-out period, and the portion of the pre-time-out period during which a response would shut off the experiment were manipulated so as to minimize the number of time outs. This was done by increasing the bird's weight, reducing the length of the pre-time-out stimulus, and reducing the length of the period at the end of the pre-time-out stimulus until the rate of responding in the pre-time-out stimulus became zero. Then, the original values were restored: 80% body weight, and a 30-second pre-time-out stimulus, with responses during the last five seconds shutting off the experiment. When suppression of responding during the pre-time-out stimulus was complete, the original pre-time-out conditions were reinstated; that is, the first response 30 seconds after the appearance of the pre-time-out stimulus shut off the apparatus for 10 minutes.

The start of the pre-time-out stimulus is indicated on the records by the diagonal movement of the recording pen. Reinforcements are not indicated.

Results

The first peck 30 seconds after the appearance of the blue light produces a time out.

The results confirm the earlier experi-

ment with chimpanzees (Exp. I) where the time out produced increased rates of responding during the pre-time-out stimulus.

First bird. Figure 5 shows the final performance for the first bird, after 51 hours of variable-interval reinforcement and 43 hours of reinforcement with the pre-time-out procedure. The figure contains a complete session except for two hours omitted between Records A and B and Records B and C. Without exception, the rate of responding is normal during the pre-time-out stimulus, even though long pauses and low rates of responding are common elsewhere, particularly after a time out, for example, at *e*. The magnitude of the pre-time-out rate of responding varies considerably. At *a* and *c* it is much below the prevailing rate. The average rate of responding is considerably higher during the blue light than elsewhere, because responding is always sustained during the pre-time-out stimulus, while pausing occurs frequently elsewhere.

Second bird. The second bird was only briefly exposed to the pre-time-out procedure (10 hours), and the responding during the pre-time-out stimulus was incompletely developed.

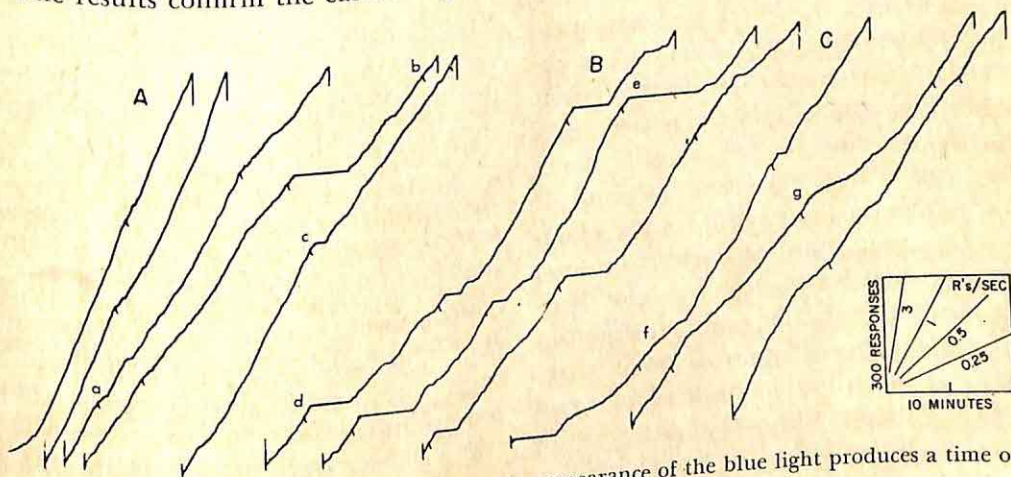


FIG. 5. First bird. The first peck 30 seconds after the appearance of the blue light produces a time out.

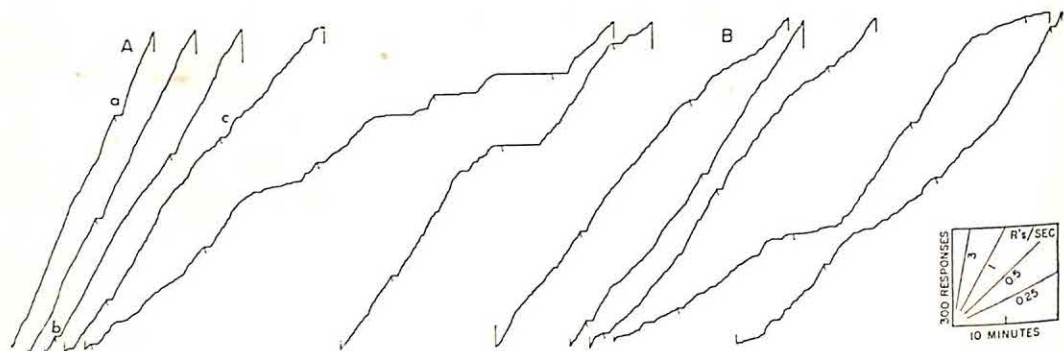


FIG. 6. First bird. Final performance showing suppression of responding during the blue light. Only responses during the final 5 seconds of the pre-time-out stimulus produce time outs.

Only responses during the latter part of the pre-time-out stimulus procedure produced a time out.

When the pre-time-out stimulus was allowed to terminate without a time out if the rate of responding were low enough, responding was almost completely suppressed during the pre-time-out stimulus.

First bird. Figure 6 shows the final performance on a 30-second pre-time-out stimulus, with standard food deprivation and responses during the last five seconds producing a time out. This performance was obtained after the following history: 14 hours in which the bird's weight was increased and the pre-time-out stimulus reduced to 15 seconds, with only responses during the latter 3 seconds producing time outs; 24 hours of an 18-second pre-time-out stimulus, with only responses during the last 5 seconds producing time outs. A period of 8 hours was omitted between Records A and B. In almost every case, suppression of responding was complete during the pre-time-out stimulus, as at *a*, *b*, and *c*. During the entire session, only 8 responses occurred during 78 presentations of the pre-time-out stimulus, and all of these occurred during the early part of the pre-time-out stimulus where they do not produce time outs.

Second bird. The second bird also showed complete suppression during the 30-second pre-time-out period. During 37 presentations of the blue light, only 9 responses occurred, but none during the period when they would produce a time out. In most cases, responding began immediately following the termination of the blue light, although the over-all rate of responding occasionally fell to zero with several successive presentations of the pre-time-out stimulus without any responding in between. The bird's history included 16 hours on this procedure in which it made no progress toward a suppression of responding during the pre-time-out stimulus. For 43 hours the weight of the bird was raised and the length of the pre-time-out stimulus then reduced to 12 seconds, with responses during the last 3 seconds producing a time out.

Return to the original pre-time-out procedure; the first response after 30 seconds produces a time out.

The return to the original time-out procedure again produced substantial rates of responding during the pre-time-out stimulus.

First bird. Record A of Fig. 7 is the start of the session following Fig. 6. Suppression was complete during the first

pre-time-out stimulus, at *a*. At the arrow, the procedure was changed so that the first response 30 seconds after the appearance of the blue light produced a time out. The blue light caused a zero rate, and no responses occurred during the remaining 16 hours of the session. The following session, Record B, began with the blue light on, and sporadic responding during the pre-time-out stimulus produced a time out after a pause of a few seconds. Records B and C show the remainder of the session, with 3 hours omitted in between. Responding was sustained during the pre-time-out stimulus with frequent instances of a rate higher than the prevailing rate, as at *c*, *d*, *e*, *g*, and *h*. The effect is somewhat more marked than that occurring during the final performance in Fig. 6. Pauses occasionally followed the time out, as in the segment at *f*, which also shows some compensation.

Second bird. The second bird's transition was similar. When the blue light remained on instead of terminating after 30 seconds, no responses occurred for almost 20 minutes, and the pre-time-out stimulus was turned off, ad libitum, with-

out a response or a time out. Responding began soon thereafter, and the next two presentations of the pre-time-out stimulus showed complete suppression, with some intervening responding. The following session began with complete suppression during the pre-time-out stimulus; and following the third presentation of the pre-time-out stimulus, the procedure was changed so that the first response after 30 seconds produced a time out. The blue light remained on for 8 hours, when no responses occurred. The following session began with the pre-time-out stimulus in effect, and a single response during the blue-light stimulus produced a time out. Thereafter, the rate of responding during the pre-time-out stimulus increased, reaching values toward the end of the session which were roughly equal to the prevailing local rate. After further exposure to the pre-time-out procedure, responding was sustained during the pre-time-out stimulus, with frequent instances of high rates of responding. Rates of responding lower than the prevailing rate continued to occur, particularly toward the beginning of the session. The rate of responding

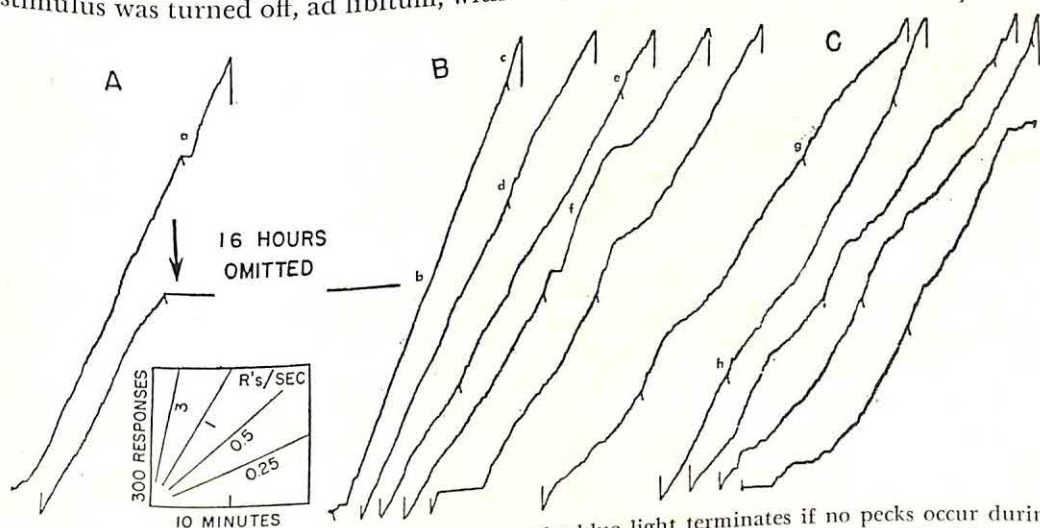


FIG. 7. First bird. Record A: Before the arrow, the blue light terminates if no pecks occur during the last 5 seconds. At the arrow, the first response 30 seconds after the appearance of the blue light interrupts the experiment. Records B and C: The session following Record A. The first response 30 seconds after the blue light produces a time out.

during the pre-time-out stimulus was approximately equal to the prevailing rate, and the performance was approximately the same as that which occurred earlier under the same procedure.

Chimpanzees

Procedure

The procedure reported for pigeons was carried out with chimpanzees No. 67 and 97. At the start of the experiment, the over-all rate of responding was low because of punishment of fast responses in a previous experiment. The pre-time-out stimulus (red light) was then presented every eight minutes. The first response 30 seconds after the appearance of the red light produced a time out. After 12 sessions for No. 97, and 9 sessions for No. 67, the rate of responding during the red light was approximately equal to the over-all rate. The procedure was then changed to emphasize the aversive properties of the time out by allowing the red light to terminate without a time out if the rate of responding were low during its latter portion. As with the pigeons, the length of the pre-time-out stimulus and the portion of the stimulus in which a response would produce a time out were varied so as to minimize the number of time outs. When the rate of responding during the red light became low enough so that nearly all of the pre-time-out stimuli terminated without a time out, the original values were reinstated.

Results

Suppression of lever pressing during the pre-time-out stimulus. Number 97 showed suppression during the pre-time-out stimulus after the following history: six sessions of a 20-second, pre-time-out stimulus with responses during the last five seconds producing time outs; three sessions where responses produced time outs during the last two seconds; and two sessions where responses produced time outs during the last five seconds. The length of the pre-time-out stimulus was then increased to 30 seconds, and Record A of Fig. 8 contains the entire second session. The over-all rate of responding was still lower than in Exp. I, II, and IV, probably because high rates

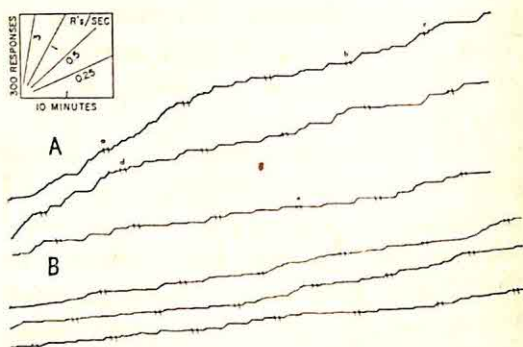


FIG. 8. Record A: Chimpanzee No. 97. Final performance. Responses during the final 5 seconds of the 30-second red light produce a time out. The oblique marks indicate the extent of the pre-time-out stimulus. Record B: Chimpanzee No. 67.

of responding were punished in a previous procedure. Suppression of lever pressing during the red light was virtually complete, with only one time out occurring (at *c*) and only 19 pre-time-out responses through the entire session. A few responses appeared in the middle of the interval, at *b*, and at the start of the interval, at *d*. The pre-time-out stimulus occasionally interrupted a substantial rate of responding, as at *a*; but the low over-all rate provided many instances where the pre-time-out stimulus occurred when the rate of responding was zero, as at *e*.

As in Exp. I, II, and IV, No. 67 showed greater sensitivity to the effects of the time out. By the second session of a 20-second pre-time-out stimulus, with responses in the final five seconds producing a time out, suppression of lever pressing in the presence of the red light was complete. No time outs occurred during the entire session, and only two responses occurred during the pre-time-out stimulus. Record B of Fig. 8 contains an entire session, showing suppression during a 30-second pre-time-out stimulus after a history the same as that of No. 97. The

over-all rate of responding was low, as with No. 97, probably because of the previous history of punishment of fast responding.

The first response after the 30-second red light produces a time out. Even after prolonged exposure, this procedure did not produce as sustained a rate of responding during the pre-time-out stimulus as during the previous performance on the same program.

The transition was very similar to that after the same procedures in the pigeon experiment, although not so extreme. Long pauses appear in the pre-time-out stimulus and elsewhere, with a corresponding decline in the over-all rate. When responding occurred, however, it occurred at normal rates of responding. No. 67, the animal with the low rate and greater reactivity to all of the previous time-out procedures, again exhibited a larger effect of the change in procedure. The over-all rate of responding fell to near zero for the last three hours of the session. Only 36 of the possible 58 reinforcements were delivered. The procedure was continued for 22 sessions, during which both the over-all rate and the rate during the pre-time-out stimulus rose; but the pre-time-out rates of responding still did not exceed the prevail-

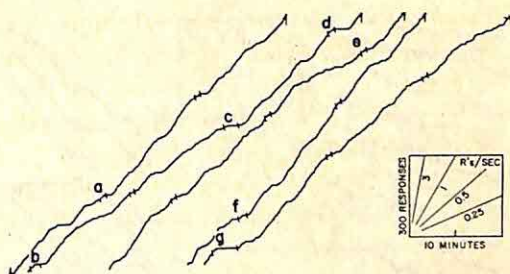


FIG. 9. Chimpanzee No. 97. The first response after 30 seconds in the red light produces a time out. Twenty-two sessions after the suppression during the previous procedure.

ing rates of responding. Figures 9 and 10 show the final performances. The over-all rates of responding remained high for both animals, and the frequency of negatively accelerated rate changes during the red light increased, as, for example, at *a, b, c, d, e, f,* and *g* in Fig. 9 and *a, b, c, d,* and *e* in Fig. 10. Occasionally, No. 67's rate of responding during the red light was near zero, as at *f* and *g*. Figure 11 shows the performances for both animals during the red light with less reduction. The segments of the curve in the vicinity of the pre-time-out stimulus are excerpted and arranged in order of occurrence, in Record A for No. 67 and Record B for No. 97.

Only responses during the last five seconds of the red light produce a time out.

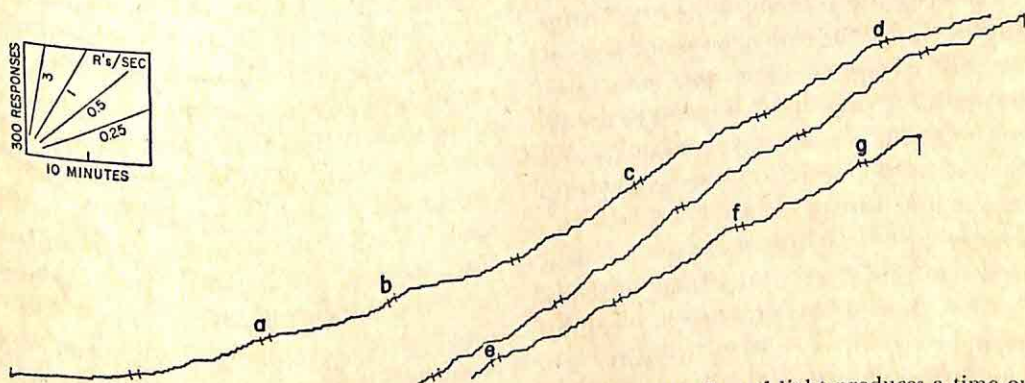


FIG. 10. Chimpanzee No. 67. The first response after 30 seconds in the red light produces a time out. Twenty sessions after the suppression during the previous procedure.

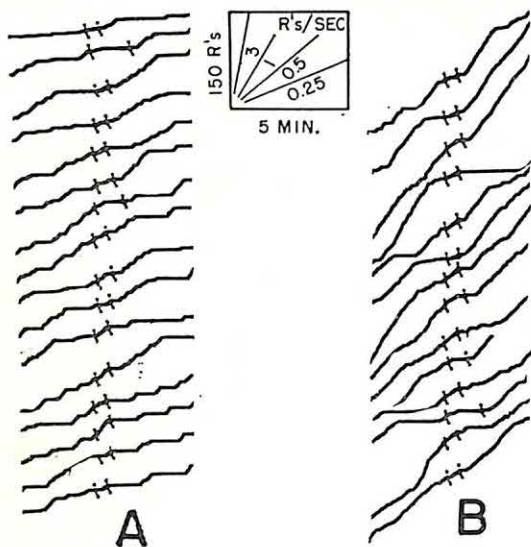


FIG. 11. Excerpts showing the performance during the red light with less reduction. Record A: Chimpanzee No. 67. Record B: Chimpanzee No. 97.

Neither No. 67 nor No. 97 progressed further toward sustained responding during the pre-time-out stimulus than the performance already described in Fig. 9 and 10. The procedure was therefore changed to find out the extent to which the earlier performance could be recovered when the pre-time-out stimulus could terminate without a time out if the rate of responding were low enough. The pre-time-out stimulus was reduced to 15 seconds, with only responses during the last 2 seconds producing time outs. After three sessions, the rate of responding during the red light became low enough so that no time outs occurred, and that portion of the pre-time-out stimulus was increased in which a response would produce a time out.

Number 67 again showed greater sensitivity to the effects of the time out. Suppression during the pre-time-out stimulus was virtually complete by three sessions, with only 34 responses occurring during the 18 presentations of the pre-

time-out stimulus. The bulk of these occurred during one pre-time-out period. The performance for No. 97 is shown in Fig. 12, where segments have been excerpted in the vicinity of the pre-time-out stimulus and arranged vertically in order of occurrence. In spite of the large amount of responding during the pre-time-out stimulus (972 responses), only one time out occurs, because most of the responding is concentrated during the early part. Marked negative acceleration occurs in all of the pre-time-out periods except the 3rd, 4th, 5th, and 11th.

Discussion

Both the pigeon and chimpanzee experiments demonstrate that the time out is an aversive event which can suppress behavior during a pre-time-out stimulus. In the pigeon experiments, it was possi-

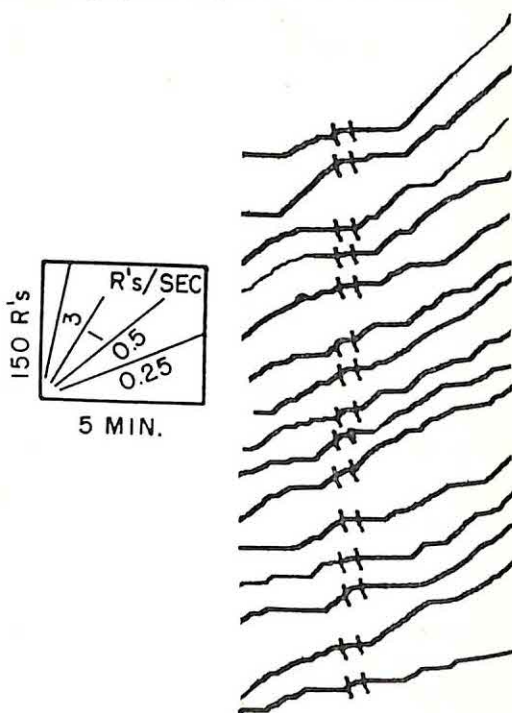


FIG. 12. Chimpanzee No. 67. Return to the procedure where only responses during the last 5 seconds of the pre-time-out stimulus produce a time out. Segments in the vicinity of the pre-time-out stimulus have been excerpted.

ble, in the same birds, to alternate between procedures which produced suppression or sustained responding at rates higher than the over-all rate. In the chimpanzee experiment, the same animals which showed sustained responding at rates higher than the over-all rate during the pre-time-out stimulus showed complete suppression when the pre-time-out stimulus could terminate without a time out. The subsequent return to the first procedure, however, did not increase the pre-time-out rates of responding to the earlier values (presumably as a function of the intervening exposure to the procedure emphasizing the aversive aspect of the time out). The increased rate during the pre-time-out stimulus is due to its effect on the rate of responding at the moment of reinforcement and the consequent differential reinforcement of high rates.

The suppression during the pre-time-out stimulus demonstrates a process in the control by a time out, which is similar to the suppression obtained with electric shock. The time out appears to have slightly different properties than the electric shock in that it has more specific action; that is, it limits the suppression to the pre-time-out stimulus. The preshock procedure often produces suppression elsewhere than during the preaversive stimulus, for example (1, 7).

The pre-time-out stimulus showed a high order of control over the pigeons' behavior, with suppression during 8 and 16 hours, respectively, for the two animals of the experiment. This order of stimulus control occurs in experiments with electric shock only at high intensities of the shock. Following the experiment producing suppression in the pre-time-out stimulus, both chimpanzees showed many instances of negatively accelerated rate changes in the pre-time-

out stimulus under the procedure that previously produced sustained responding. The suppressive effect of the time out during the final part of the pre-time-out stimulus probably represents the concurrent operation of the two properties of the time out; the sustained responding early in the pre-time-out stimulus is due to the differential reinforcement of high rates by the time out, while the lower rate toward the end of the pre-time-out stimulus represents a temporal discrimination of the aversive effects of the time out. When only responses during the latter part of the pre-time-out stimulus produced time outs, the temporal discrimination became more pronounced and the complete suppression that developed during the first exposure to this procedure did not appear. This is similar to the result obtained by Azrin, who generated near-normal rates of responding during the first part of a preshock stimulus.

The pigeons showed no indication that a temporal discrimination was developing, but they were not exposed to the procedure as long as the chimpanzees.

EXPERIMENT IV. AVOIDANCE BEHAVIOR

Another property of an aversive stimulus is its ability to maintain behavior which removes or postpones it. Sidman has extensively studied the avoidance of electric shock, with procedures in which a free operant is maintained because it postpones a shock. Morse and Herrnstein (13) extended the technique to the avoidance of a time out in the pigeon; and in the present experiment, this technique is used with the chimpanzee. The demonstration of the aversive properties of the time out is relevant to the interpretation of the results of Exp. I and III, where findings were analyzed by reference to two separate functions of the

time out. The animals, the schedule of food reinforcement, and the level of deprivation remain the same. Hence, the aversive properties of the time out derived from the present experiment are germane to the interpretation of the functions of the time out in those experiments.

Procedure

The experimental session began with the variable-interval schedule in force. Two keys separated by approximately 12 inches were present simultaneously. Responses on the right key (food key) were reinforced on the three-minute, variable-interval schedule used in the previous experiment. If no responses occurred on the left key (avoidance key), a three-minute time out occurred after every 45 seconds of experimental time. Each time the avoidance key was pressed it postponed the interruption of the experiment for a given period of time (RS interval). If the avoidance key was pressed frequently enough, the variable-interval schedule remained in force continuously. Four values of the RS interval were studied: 10, 5, 2, and 1 minutes. Figure 13 shows the actual sequence and the number of sessions in each value for both animals.

Since the food reinforcement was programmed by time, a sequence of responses on the avoidance key would increase the probability that the next response on the food key would be reinforced. Thus, the rate of avoidance responding would be spuriously inflated (12, 17.) Such accidental reinforcement was observed when the rate of No. 97's avoidance responding failed to fall following the increase of RS interval from 2 to 10 minutes. To prevent this spurious reinforcement, the timer programming the food-reinforcement schedule stopped for seven seconds every time an avoidance response occurred. This was done with both animals. This procedure increased the mean of the variable-interval schedule somewhat; but it ensured that the level of avoidance responding would not be inflated by reinforcement because of a spurious chain of responses.

The avoidance experiment began following the procedures already described in Exp. I and II. No responses occurred on the avoidance key for several sessions, however; and, because some responding was necessary before the avoidance contingency would have any effect, the food key was withdrawn for 30 minutes during which avoidance-key responses were reinforced by food on the variable-interval schedule. The food key was then replaced and the avoidance procedure re-instated.

Responding on the food and avoidance keys was recorded on separate cumulative recorders running concurrently. The displacement of the recording pen diagonally from the curve marked the occurrence of food reinforcement on the variable-interval curve and time outs on the avoidance curve. The recorder stopped during the time out, and the location of the time outs on the variable-interval curve is indicated by the small filled circles shown above the curve in Fig. 16. Figures of the avoidance behavior contain entire daily sessions.

Results

The results of the entire experiment are summarized in Fig. 13, which gives daily rates of responding for each session on the food and avoidance keys, as well as the frequency of occurrence of time outs. All of the sessions of the experiment are included, and each point is the average rate for the daily session. In general, the rate of avoidance responding increased when the interval by which each response postpones the time out was decreased, and vice versa. On all values of the avoidance schedule, rates of responding were considerably higher than the minimum amount of behavior necessary to postpone all of the time outs. Responding tended to occur in bursts at sustained rates rather than in single responses emitted at approximately the rate corresponding to the avoidance interval.

Chimpanzee No. 67. Figure 14 gives the early development of the avoidance behavior. Record A, the sixth session in which each avoidance response postponed the time out for 10 minutes, shows a very low rate of responding—one which is just sufficient to keep the schedule of food reinforcement in force continuously during most of the session. Seventy responses are emitted during the session, in bursts of 2 to 10. Record B contains the performance in the procedure where avoidance responses postponed the time out for 5 minutes (RS 5). It shows a

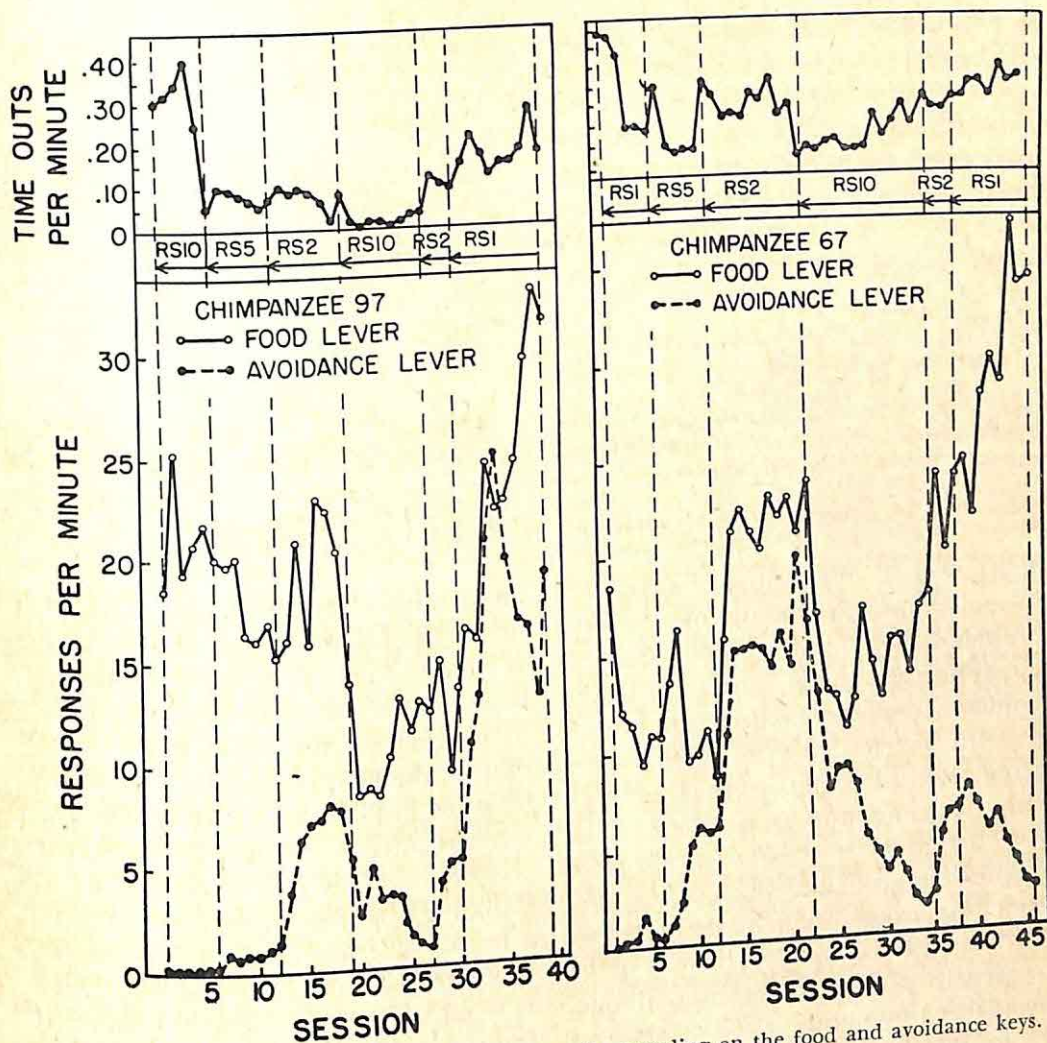


FIG. 13. Chimpanzees No. 67 and 97. Daily rates of responding on the food and avoidance keys.

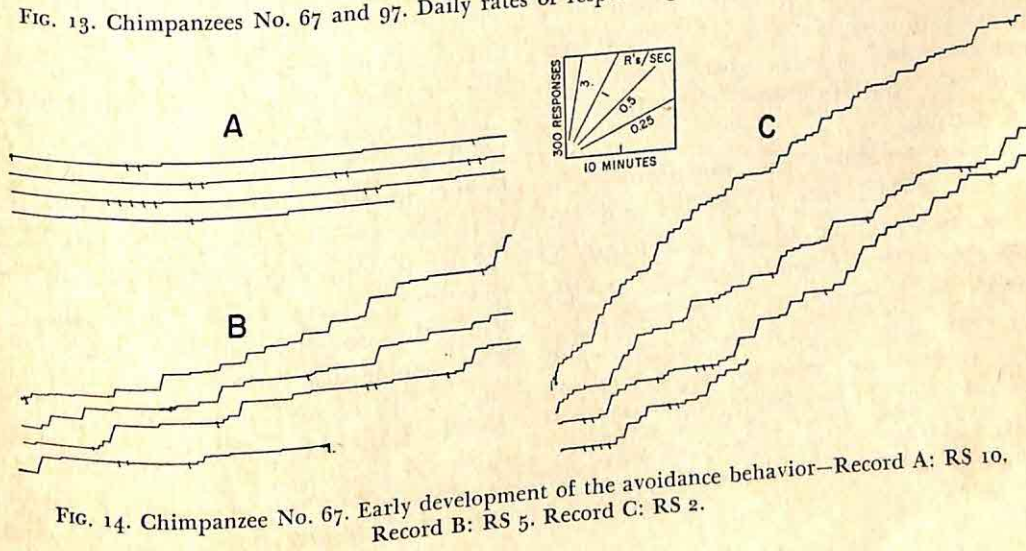


FIG. 14. Chimpanzee No. 67. Early development of the avoidance behavior—Record A: RS 10. Record B: RS 5. Record C: RS 2.

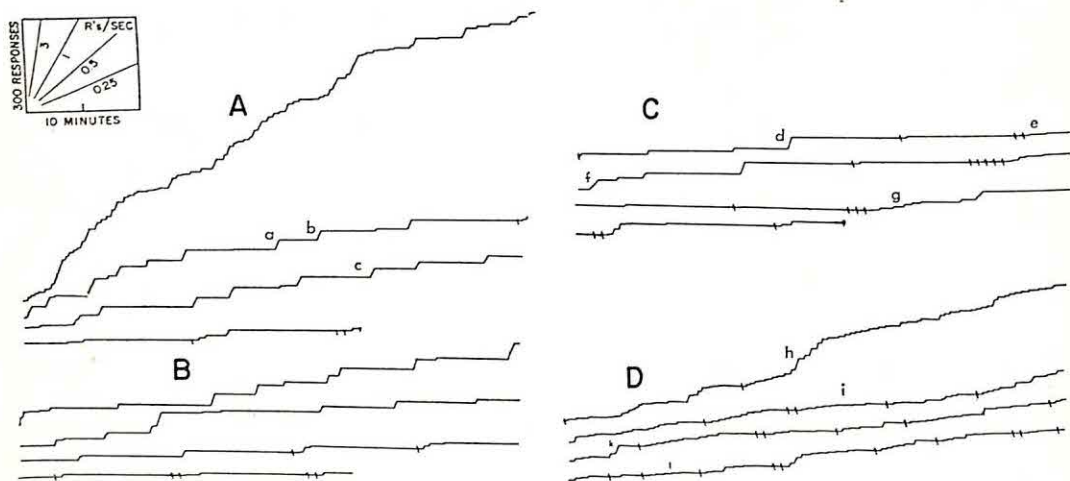


FIG. 15. Chimpanzee No. 67. Transition to RS 10 and subsequent return to RS 2.

higher over-all rate than that under RS 10 and a progressive decline in rate through the session. Intermediate rates of responding are absent, with most of the avoidance responses occurring in groups of 20 to 60, emitted at 2 responses per second. Record C contains the avoidance performance after 9 sessions on RS 2. The over-all rate increases over twice that of the previous procedure, but a larger number of time outs occur, particularly toward the end of the session, because of the progressive decline in the over-all rate. Avoidance responding is absent for periods up to 5 minutes, during which three to four time outs occur in a row. Intermediate rates of responding are still absent, although the size of the bursts is decreasing.

Figure 15 shows the transition of the avoidance behavior to RS 10 and the subsequent return to RS 2. Record A, containing the session following Fig. 14C, shows the re-exposure to the longer RS interval. The session begins with the high rate of avoidance responding from the previous shorter RS interval; but the rate of responding falls progressively, reaching a low over-all rate of respond-

ing consisting of long pauses separated by bursts of approximately 25 responses, as at *a*, *b*, and *c*. Eight sessions later, shown in Record B, the rate of avoidance responding remains substantially higher than in the final performance previously recorded in Fig. 14C. This level had been reached after the fourth session on RS 10. Because the rate of avoidance responding might be spuriously inflated as a result of accidental reinforcement, the procedure was then changed so that the tape scheduling the variable-interval reinforcement stopped for seven seconds each time an avoidance response occurred. Thus, responses on the avoidance key no longer increased the likelihood that subsequent responding on the variable-interval key would produce a reinforcement. Record C shows the performance four sessions after the initiation of the control procedure. The rate of responding falls to a level only slightly above the previous final performance on RS 10. Some sustained bursts of responding are still occurring, as at *d*, but intermediate rates of responding have become much more frequent, as at *e*, *f*, and *g*. The number of time outs is now com-

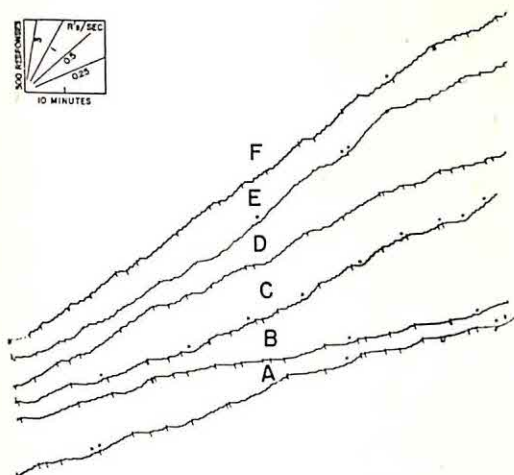
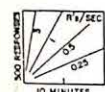


FIG. 16. Chimpanzee No. 67. Food-key performances corresponding to the avoidance behavior of Fig. 14 and 15.

parable with that recorded under the previous RS 10 procedure. The RS interval was then reduced to two minutes; Record D shows the performance after three sessions on this value. The rate of responding has increased considerably over the performance recorded previously under RS 10 in Record C; but it is much less than the performance recorded earlier under RS 2 in Fig. 14. This is largely a result of the disappearance of the sustained bursts of responding at high rates. Instead, the perform-

ance has given way to a finer grain in which many responses are occurring singly, or in groups of two or three, as at *j* and *l*. Instances still occur, however, where avoidance responding is sustained at high rates for large numbers of responses, as at *h* and *k*.

The corresponding performances on the food key are shown in Fig. 16, which gives the first hour of the session. The order of presentation of the segments has been inverted for more compact presentation. Records A and B are the food-key performances, comparable with Fig. 14 and 15 on RS 10. Record C shows the increase in the variable-interval rate of responding on the first session the RS interval was increased to two minutes. Records D and E contain the 8th and 12th sessions on the second exposure to the RS 10, and Record F shows the subsequent increase in rate of responding on re-exposure to RS 2 in the final session.

When the RS interval was decreased to one minute, the rate of responding began to decline continuously, finally reaching a value intermediate between the lowest RS 10 value and the performance under RS 5. Figure 17 shows the second through

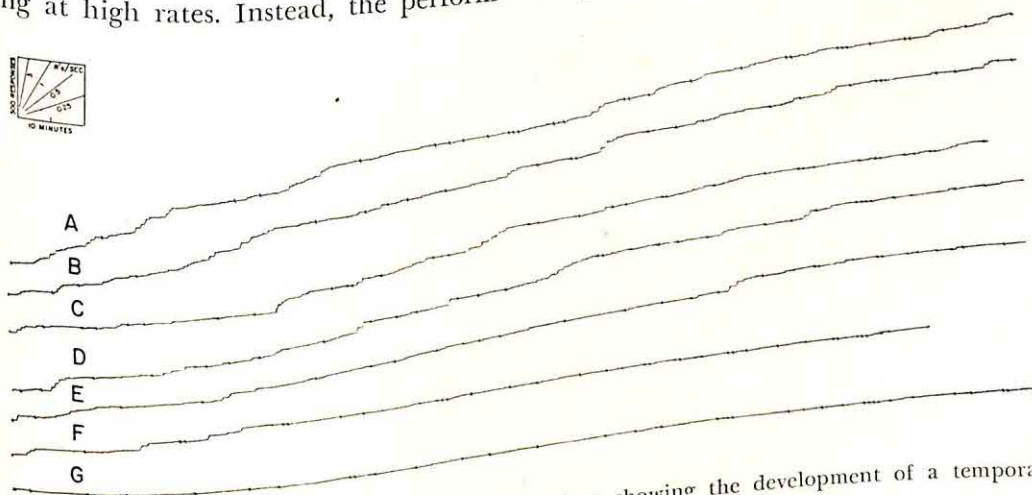


FIG. 17. Chimpanzee No. 67. Seven consecutive sessions showing the development of a temporal discrimination in the avoidance responding.

the eighth sessions in reduced form. The over-all rate of responding declines continuously, because of the progressive disappearance of bursts of responding at high rates. By Record G, responding is sustained at a low constant rate, just sufficient to postpone most of the time outs. The frequency of time outs is higher toward the end of the session.

Figure 18 shows the final performance in greater detail. The avoidance behavior in Record B occurs at a low constant rate of responding, just sufficient to postpone most of the time outs. A brief burst of responding at a high rate occurs at *a*. The variable-interval performance on the food key, Record A, shows the increase in the over-all variable-interval rate. The small arrows over the curve indicate the exact locations of the time outs. Most of the time outs occur while the variable-interval performance is sustained, except for the group of eight time outs toward the end of the sessions, which occur during a pause. Nearly all of the time outs are followed by a brief pause in the variable-interval performance.

Throughout the experiment the over-all rate of responding on the food key went up whenever the number of time outs increased and decreased whenever the number of time outs fell. This relation confirms the findings of Exp. I and II on the effect of an ad libitum time out

on a variable-interval performance.

Chimpanzee No. 97. The second animal in the experiment did not show so strong a disposition to emit large bursts of responding on the avoidance key during the early parts of the experiment. However, the avoidance rates of responding increased markedly under RS 1 toward the end of the experiment. Some evidence existed that temporal discrimination had started to form under RS 1, as with No. 67, but the experiment was discontinued too soon to show its development.

As in the previous experiments, No. 97 was less sensitive to the aversive effects of the time out.

Discussion

The present experiment confirms Morse and Herrnstein's result with the pigeon and shows that the postponement of a 10-minute time out from a variable-interval schedule of food reinforcement will maintain a substantial amount of behavior in a chimpanzee. The results also extend the generality of the aversive properties of the time out from this schedule of reinforcement, and support the interpretation of the results of Exp. II.

The avoidance behavior based on time out from positive reinforcement has some of the properties of avoidance of electric shock studied extensively by Sid-

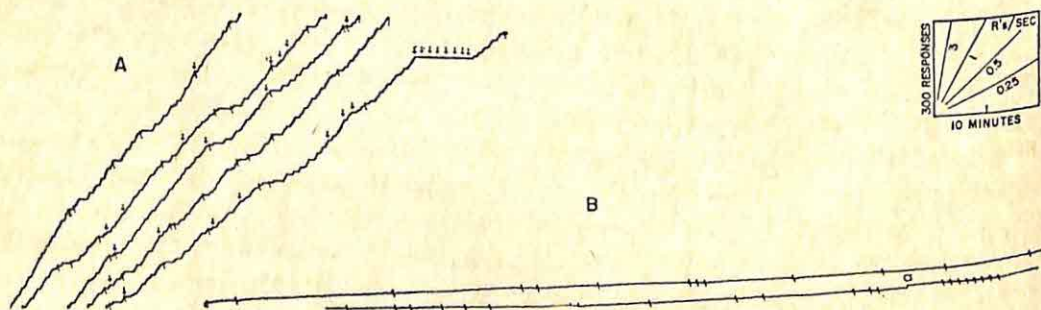


FIG. 18. Chimpanzee No. 67. Final performance showing full development of a temporal discrimination.

man. Over most of the range studied in this experiment the rate of avoidance responding varied inversely with the RS interval (15). With prolonged exposure to the avoidance contingency, avoidance rates tended to be progressively lower because of the development of a temporal discrimination. The initial rate of development of the avoidance behavior was similar to the results with electric shock in the rat, cat, and monkey, although the 10-minute RS interval in the present experiment was considerably larger than the values customarily used.

The high chimpanzee avoidance rate was characterized by bursts of responses separated by pauses, rather than the steadier intermediate rates of responding of the rat, cat, and monkey. Immediately following a time out, the rate of responding of the chimpanzee is uniformly low. With the rat, cat, and monkey experiments, however, shocks tend to be followed by a burst of avoidance responding. Because species properties are confounded in the relevant experiments, it is impossible to say yet what properties of avoidance behavior based on a noxious stimulus are different from those of avoidance behavior based on time out from a schedule of positive reinforcement. The large amount of uniformity in the properties of electric-shock avoidance over a wide range of vertebrates suggests that the important condition is the use of the time out rather than the species difference. Avoidance of time out from positive reinforcement has not been studied in other species extensively enough so that the role of the species in this behavior can be assessed. Experiments on the pigeon in which the time out was used as a punishment (see Exp. V) suggest that the pigeon is less sensitive to this contingency than the chimpanzee. The progressive increase in the rate of

responding on the food key during the last exposure to RS 10, 2, and 1 minutes occurs independently of changes in the avoidance rate. The magnitude of the increase of the food-key rates is greater than the change in the avoidance rates, and cannot be due to a simple shift of responses from the avoidance key to the food key. Nor is the increased food rate due to the periodically recurring time outs, which can of themselves produce a marked increase in rate. Earlier in the experiment, equal frequencies of time outs resulted in a much lower food-key rate. The increase in the food-key rate is probably related to the development of the avoidance behavior, and may be related to the temporal discrimination which developed concurrently on the avoidance key.

Avoidance of a time out from a schedule of reinforcement raises some questions about the basis of avoidance behavior. Most writers discussing avoidance of electric shock interpret the performance in terms of Pavlovian conditioning, based on reflexes elicited by the shock. Some feel that the reduction in these reflexes strengthens the avoidance behavior (10, 11). Others emphasize the punishment of all behaviors other than the avoidance response with the consequent development of these behaviors as conditioned stimuli for the reflex effects of the shock (4, 14, 15). Even though a time out from a schedule of positive reinforcement undoubtedly has some emotional effects, these probably are not of the magnitude caused by a high-voltage electric shock. Yet, avoidance behavior can still be maintained at a level comparable with that developed using shock. This suggests that an aversive stimulus may control behavior independently of its emotional, autonomic, or reflex effect.

EXPERIMENT V. TIME OUT FROM A SCHEDULE OF POSITIVE REINFORCE- MENT AS A PUNISHMENT

This experiment, using the same chimpanzees as Exp. I-IV, investigates the properties of a time out used as a punishment. A time out followed "fast responding" and reduced the rate of responding on a variable-interval schedule of reinforcement. Because the variable-interval schedule and the level of deprivation were the same as in the previous experiments, the results of this experiment, as well as the results of the earlier experiments in the paper, can be used to describe the general properties of this time out.

The variable-interval schedule was used as a paradigm for studying punishment because the constant rate of responding that emerges is not easily disrupted by the time out independently of its aversive effects, as it is in other interval schedules. Although a fixed-interval schedule, for example, has the advantages of the sensitivity and reproducibility of rate changes, the interruption of the experiment per se would disrupt the fixed-interval performance. Also, an interval schedule was used rather than a ratio schedule because of its moderate over-all rate of responding and frequent local rate changes. In a fixed-ratio schedule, with its sustained responding at a single high rate, the punishment would have little differential effect, since very few responses would go unpunished. Moreover, the schedule factors in the ratio schedule differentially reinforce high rates of responding, which oppose the effects of the punishment (7). The punishment of responses maintained on a fixed-ratio schedule is analogous to the differential reinforcement of low rates on a fixed-ratio schedule.

Only occasional instances of responses that qualified for punishment were actually punished. Therefore, the frequency of occurrence of the punished class of responses would be minimally constrained and could vary continuously over a wide range.

In the paradigm chosen here to study punishment, all responses are reinforced, and the punishment is used to reduce the frequency of occurrence of a selected class of responses. Another paradigm is possible in which a time out is used to eliminate behaviors which are occurring by induction, as, for example, in an extinction component of a multiple schedule, or in the differentiation of the topography of some behavior. These are important cases, but the rate of responding under the variable-interval schedule seemed to be a better case to study first because it permits a return to a base-line condition.

For convenience, I shall adopt the convention of designating the time between a given response and the preceding response as an interresponse time (IRT) and refer to the punishment of interresponse times even though the correct statement would be the punishment of a response having an interresponse time of a given value.

Procedure

The experiment was carried out following the procedures described in Exp. I through IV. Under the punishment procedure, lever presses spaced less than a given interval from a preceding response were followed by a time out. Every such instance, however, was not punished. The punishment occurred on a variable-interval schedule with a mean of 4.6 minutes and a range of 9 minutes. When the variable-interval punishment schedule designated a punishment, the first instance thereafter of a response qualifying for punishment was followed by a time out.

Recording

Both cumulative records and interresponse

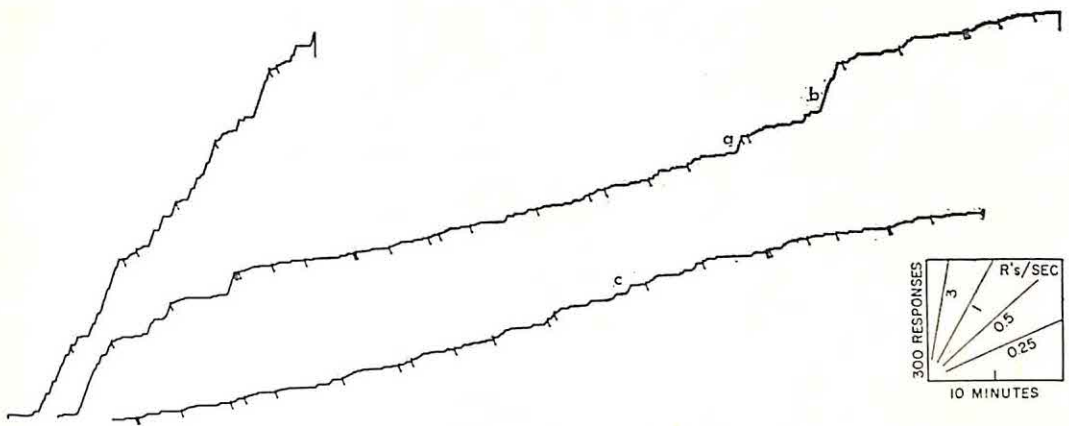


FIG. 19. Chimpanzee No. 97. Final performance on VI 3.

times, were recorded. Reinforcements are indicated by the small marks diagonal to the record, and the points of occurrence of a time out are indicated either by the small dot or arrow over the record. The scale of the record changes from figure to figure; but both the ordinate and abscissa were compressed equally, so that the slopes of the curves are comparable. The inter-response-time recorder classified all responses occurring during the session into those between zero and 0.33 second from a previous response; 0.33 to 0.66; 0.66 to 1 second; 1 to 1.33; 1.33 to 1.66; 1.66 to 2 seconds; and thence by steps of 2 seconds until 22 seconds. Because this proved to be too fine a grain, not all of these compartments were used in the final analysis. The error of the interresponse-time recorder was less than 5%.

Plan of the experiment

A variable-interval base line was first established. Responses occurring within 0.75 second from a previous response were punished as specified by the variable-interval schedule designated above. This 0.75-second interresponse time proved to be too high a value for one animal, whose rate of responding was so high that nearly all of its responses fell within the punished category. The contingency was then shifted to the punishment of responses spaced 0.33 second or less. Performances were then recorded as the over-all rate of responding was progressively lowered by punishing responses spaced 1, 2, and 4 seconds. The

time out was then removed, the variable-interval base line without punishment recovered, and the punishment procedure reinstated with responses occurring within 4 seconds of a previous response being punished. The experiment was discontinued for five weeks during the variable-interval reinforcement before the reintroduction of the punishment.

Results

Transition from No Punishment to Punishment of Responses Spaced Less Than 0.33 Second

Number 97. Figure 19 contains a final performance for chimpanzee No. 97 on the variable-interval schedule. The over-all rate of responding is high during the first 30 minutes of the session but then falls rapidly, from 0.75 response per second to less than 0.2 response per second. Local rates of responding continue to be high even toward the end of the session, as at *a*, *b*, and *c*, for example. Although this markedly negative curvature is a reproducible feature at this stage of development, it is an atypical variable-interval performance and probably a function of the previous exposure to the contingencies of Exp. I, II, and IV. On the following session, responses spaced 0.75

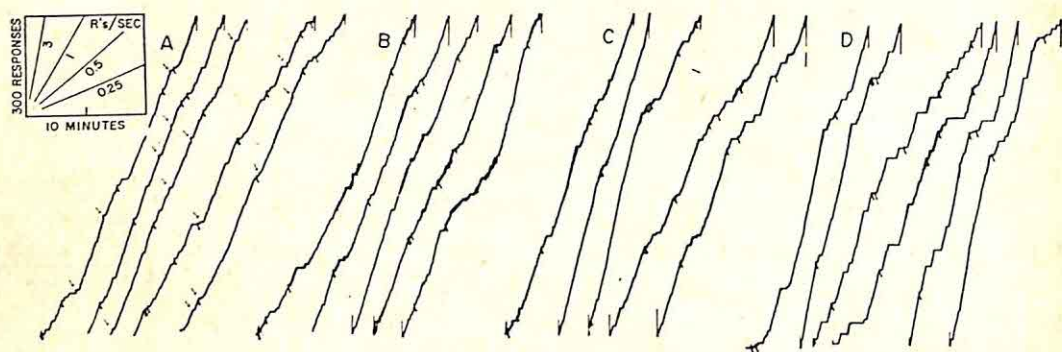


FIG. 20. Chimpanzee No. 97. The first hour of the first four sessions on the punishment procedure.

second were punished. The initial effect of the punishment is shown in Fig. 20, containing approximately the first hour of the first four sessions. In Record A the over-all rate increases from 1 to 1.5 responses per second. This performance is maintained throughout the entire session, during which 8,000 responses occurred.

Few responses are spaced *more* than 0.33 second, and the maximum number of time outs (indicated by the arrows in Record A only) occur. In the next session (Record B), 10,000 responses occurred with local rates of responding slightly less than 3 responses per second, so that nearly all of the responses in the session qualify for punishment. Beginning with the third session (Record C), responses spaced less than 0.33 second are punished. The rate of responding increases even further, reaching 10,300 responses in the third session and 12,700 responses in the fourth session (Record D), with local rates of responding reaching just under 4 responses per second.

Figure 21 contains the entire record of the third to eighth sessions in reduced form. In Record A, the rate of responding begins at the previous prevailing rate, but soon falls off to an intermediate value somewhat less than that occurring at the start of the session during the

previous variable-interval reinforcement without punishment. Despite the marked fall in the over-all rate of responding, however, the frequency of responses spaced less than 0.33 second is high enough so that the maximum number of time outs is still being received. Records B through F contain the fourth through eighth sessions, during which the over-all rate of responding falls, the negative

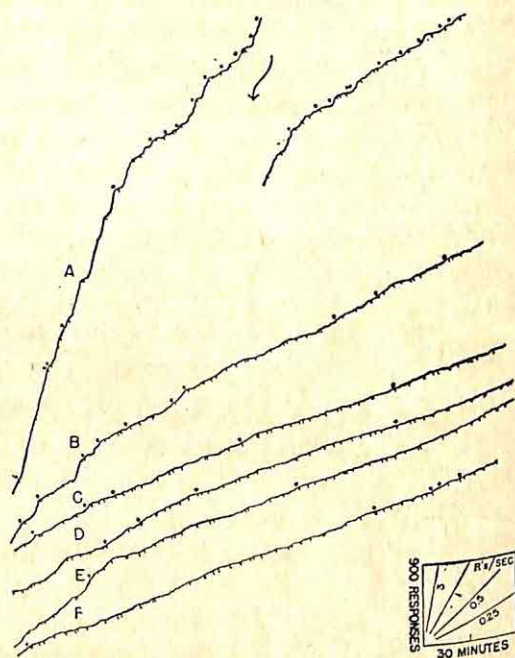


FIG. 21. Chimpanzee No. 97. Responses spaced less than 0.33 second are punished. Third to eighth sessions.

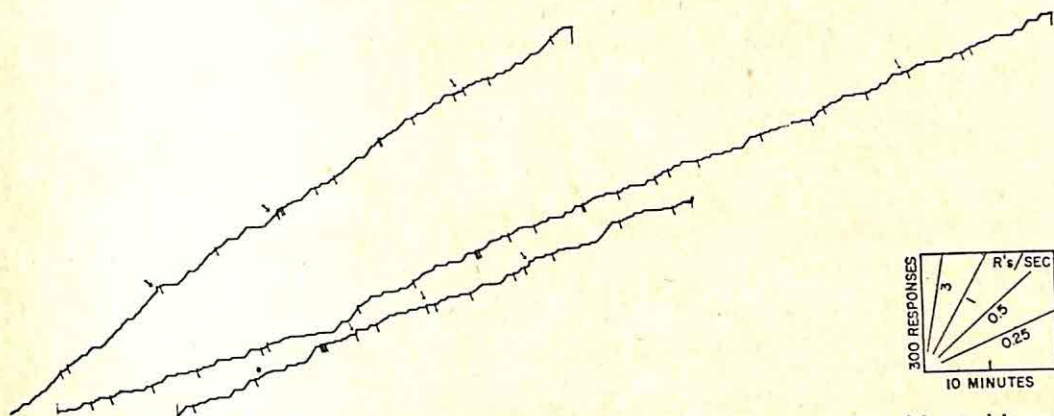


FIG. 22. Chimpanzee No. 97. Complete daily session showing the final performance with punishment of responses spaced less than 0.33 second.

curvature diminishes, and the number of time outs occurring reaches a low value.

The final performance is shown in Fig. 22, where the figure is reduced less. The number of responses emitted during the session is still the same as the number emitted before the punishment procedure (Fig. 19), even though responses spaced 0.33 second from preceding responses have been virtually eliminated (see Fig. 24). The marked negative curvature that previously prevailed has now given way to a more constant rate of responding. Some negative curvatures still exist, however, with the over-all rate falling from 0.45 response per second at the start of the session to slightly less than 0.2 response per second for the remainder of the session.

Number 67. The second subject, No. 67, again showed a greater sensitivity to the aversive effects of the time out by a progressive decline in the rate of responding as responses with larger interresponse time were punished. Records A-D of Fig. 23 show the development of the final performance with the punishment of responses spaced 0.33 second or more.

Record A contains the final perform-

ance of the variable-interval schedule without punishment. The normal variable-interval performance for this animal was characteristically lower than that for No. 97, and the characteristic negative acceleration was absent. The first two sessions (Records B and C), with punishment of responses spaced less than 0.75 second, show a progressive decline in the rate of responding, although instances of responses occurring within 0.75 second of each other still occurred, as the large number of time outs indicates. Record D contains the final performance 11 sessions after the criterion for punishment was changed to the punishment of responses spaced 0.33 second, or less. The over-all rate of responding increased somewhat, but the frequency of responses spaced 0.33 second apart, or less, was near zero, as evident from the fact that only one time out occurred.

The interresponse-time distributions of No. 67 for the prepunishment sessions (Sessions 1-7) show the generally prevailing low rate of this animal (Fig. 24). Responses spaced less than 0.33 second rarely occurred, and only 40% were spaced 1 second or less. Responses tended to occur more equally in all of the IRT categories than with No. 97. Responses

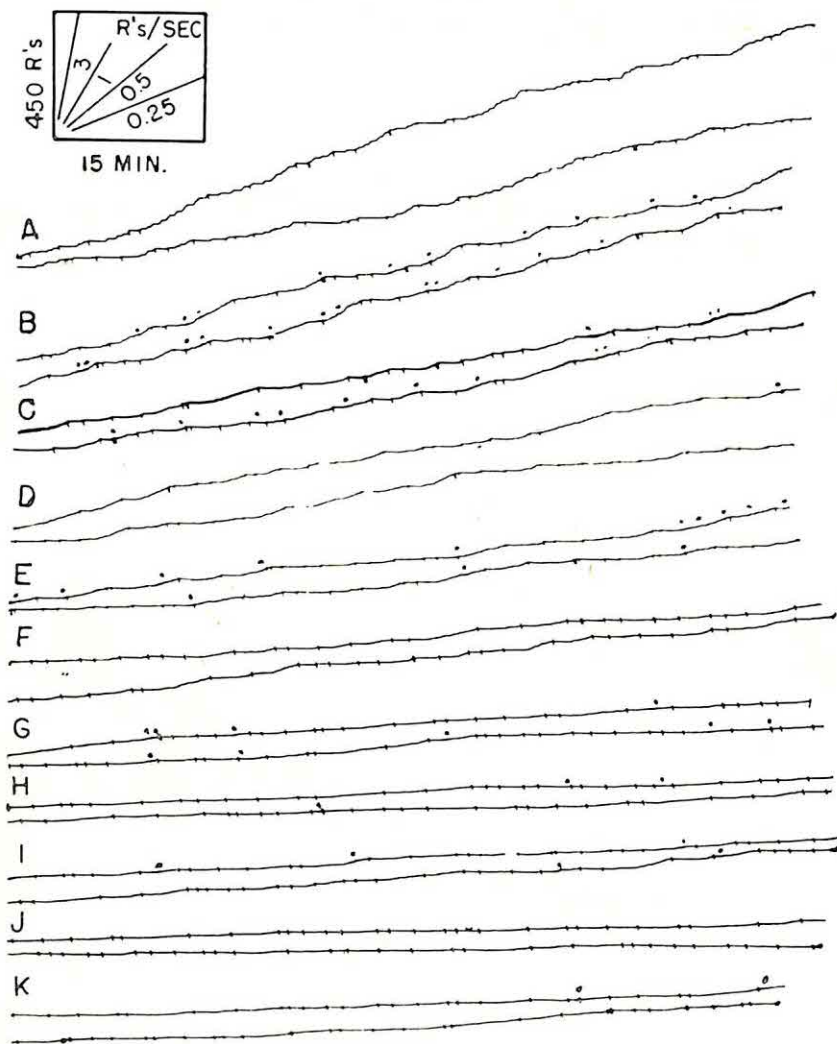


FIG. 23. Chimpanzee No. 67. Complete daily sessions showing performances with punishment of responses spaced less than 0.33, 1, 2, and 4 seconds.

A. Final variable-interval performance. No punishment.

B & C. Responses spaced less than 0.75 second are punished.

D. Responses spaced less than 0.33 second are punished.

E & F. Responses spaced less than 1 second are punished.

G, H, and I. Responses spaced less than 2 seconds are punished.

K. Responses spaced less than 4 seconds are punished.

spaced less than 1 second showed the largest decline in the frequency, when responses spaced less than 0.75 second were punished in Session 8. The remaining categories show progressively less effect of the punishment. During the remaining period of punishment of responses

spaced less than 0.33 second, the 1-second IRT's fell to near zero; the 2- and 4-second IRT's showed a marked decline; but IRT's 6 to 18 seconds leveled off at a value slightly lower than the prepunishment base line.

Figure 25 is a daily plot, for No. 97, of

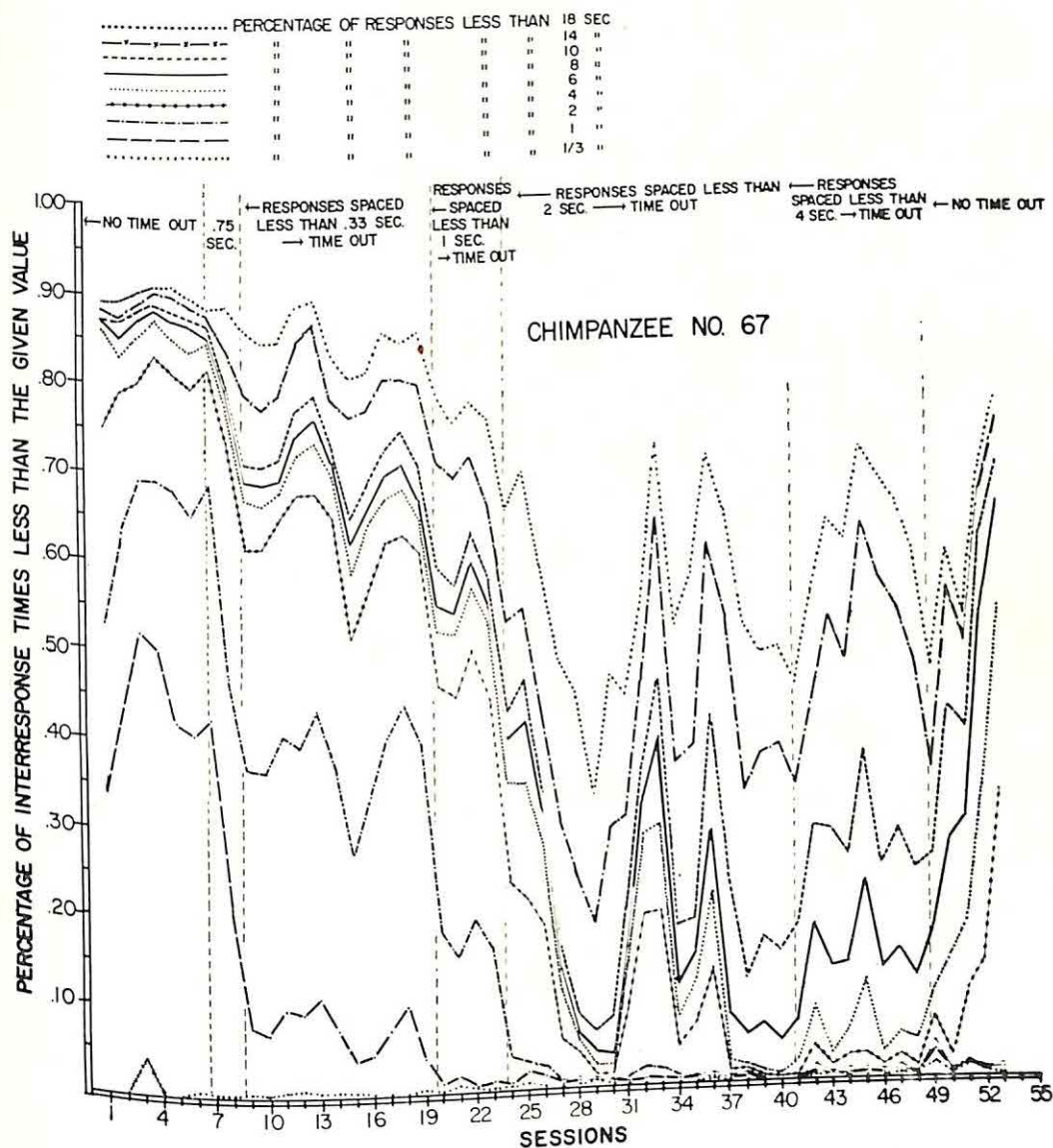


FIG. 24. Chimpanzee No. 67. Interresponse-time summary of the experiment.

the percentage of the total number of responses falling within each interresponse-time classification. The bottom curve contains the percentage of the responses in a given session which were spaced 0.33 second or less from a preceding response. The second curve contains the percentage of responses spaced 1 second, and so forth, as indicated by the figure legend. The top curve gives the percentage of re-

sponses spaced 18 seconds apart or less, and the difference between the top curve and 100% gives the percentage of responses in the daily session which were spaced 14 seconds or more apart.

The transition to the punishment of responses spaced 0.75 second or less occurs on the 10th session. Before this, 10% of the responses were spaced 0.33 second or less and most were spaced 1 second or

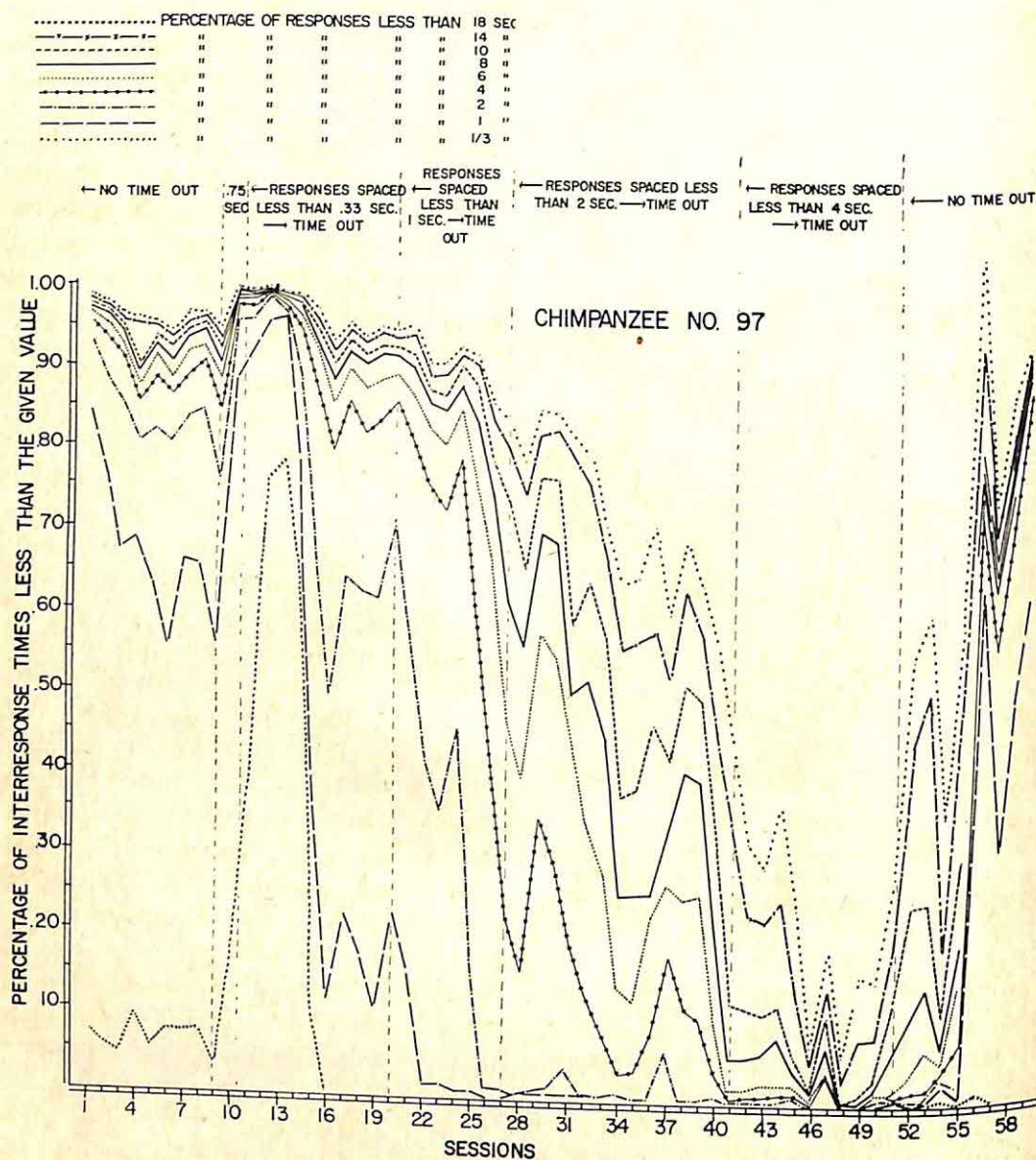


FIG. 25. Chimpanzee No. 97. Interresponse-time summary of the experiment.

less. Only 10% of the responses in this session are spaced 6 seconds apart or more. The first introduction of the punishment (Session 10, Fig. 20A) approximately triples the frequency of responses spaced 0.33 second, and almost 90% of the total responses of the session are spaced within 1 second of each other. In the second punishment session (Session 11, Fig. 20B) the rate of responding in-

creases further, with almost 90% of the responses still spaced 1 second or less. When the punishment criterion was changed to 0.33 second (in Session 12, Fig. 20C), the rate increased further, with 75% of the total responses of the session spaced less than 0.33 second and 96% spaced less than 1 second. During the subsequent decline in the rate of responding from Session 14 to 19 (Fig. 21A-E and

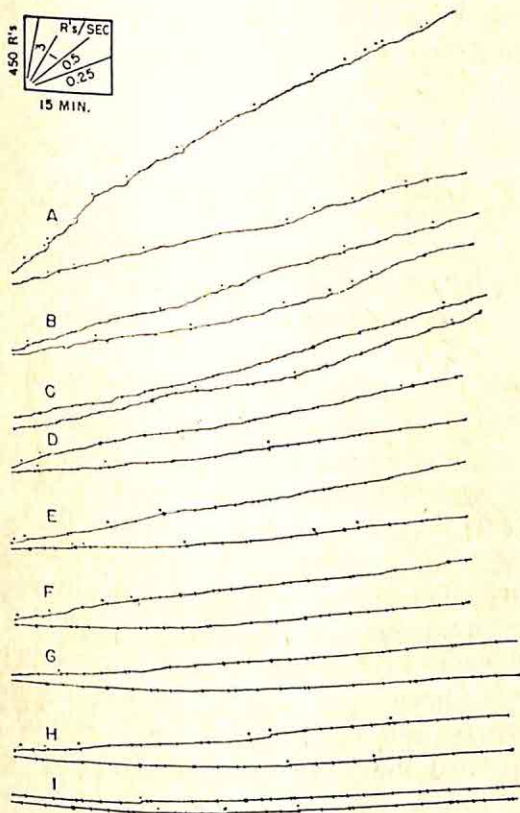


FIG. 26. Chimpanzee No. 97. Complete daily sessions showing performances with punishment of responses spaced less than 1, 2, and 4 seconds. A, B, C, and D, 1 second; E, F, and G, 2 seconds; H and I, 4 seconds.

Fig. 22), the percentage of responses spaced less than 0.33 second reached zero by the 16th session. Responses spaced more than 0.33 second also showed a decline in frequency, but the decline was progressively less marked for the lower rates. The frequencies of the various interresponse times remained stable during the remainder of the sessions.

Punishment of Responses Spaced Less Than 1 Second

Number 97. Figure 26 contains complete daily sessions for No. 97 in reduced form taken from 35 sessions during which the punishment criterion was increased to 1, 2, and 4 seconds.

Records A, B, C, and D represent the

transition to the final performance during six consecutive sessions of punishment of 1-second IRT's. Record A, containing the first session, follows Fig. 22. The over-all rate began somewhat higher, and the number of time outs increased markedly over the value prevailing in the final performance with the smaller value. By the end of the session, however, the rate of responding fell to 0.15 response per second, compared with 0.25 response per second at the end of the previous session. In the second session (Record B) the negative curvature was greatly reduced, and the over-all rate of responding remained approximately 0.2 response per second. A large number of time outs still occurred. By the third session (Record C) the over-all rate of responding fell to approximately 0.15 response per second; and even though the over-all rate of responding was considerably below 1 response per second, instances of responses spaced 1 second or less still occurred occasionally, as the four time outs occurring during the session indicate. The final performance is shown in Record D, where the over-all rate of responding fell ever further. Responding at the start of the session began at approximately 0.15 response per second and fell to less than 0.5 response per second by the middle of the session.

Number 67. Number 67's performance is shown in Records E and F of Fig. 23, containing the sessions immediately following Record D. The punishment produces an almost immediate reduction in the over-all rate of responding, along with an increase in the number of time outs. In Record F, four sessions later, the rate of responding was extremely low during the first part of the session and higher during the latter part. No instance of a response spaced less than 1 second from a previous one occurred during the session.

Punishment of Responses Spaced Less than 2 Seconds

Number 97. Records E, F, and G of Fig. 26 contain the 1st, 6th, and 11th sessions for No. 97. Record E, which follows directly from the session in Record D, contains the first exposure to the new value. The rate of responding and the number of time outs do not change, because the previous punishment of responses spaced 1 second from a previous response depressed the rate of responding sufficiently so that very few pauses as short as two seconds occurred. By the end of Record E the new contingency produced the lowest rate yet observed. Six sessions later, in Record F, the over-all rate showed little change. Four time outs occurred during the initial high rate at the start of the session. By Record G, the 11th session, the over-all rate fell even further. The negatively accelerated portion at the start of the session was much reduced, and only one time out occurred.

Number 67. Record G of Fig. 23, the session following Record F, shows the first punishment of responses spaced less than two seconds for No. 67. There is a slight decrease in the rate during the latter part of the session and an increase in the frequency of time outs. The fourth session with this value (Record H) shows an even lower rate of responding, with fewer time outs; but after a total of 13 sessions with this value (Record I), the rate of responding increased to a value approximately the same as that prevail-

ing when this value was first introduced in Record G.

Punishment of Responses Spaced Less than Four Seconds

Number 97. In Record H of Fig. 26, the session following Record G, the initial effect of the new contingency was a slight increase in the frequency of time outs and the over-all rate of responding during the first part of the session. The rate fell markedly by the end of the session, however. The lowest rate of responding on this value is shown in Record I, taken after 11 sessions. The high rate of responding that characterized the start of the session heretofore disappeared, and the performance became a constant rate throughout the session. Although the over-all rate of responding was already less than a response every four seconds, the punishment of the occasional four-second IRT's produced a further decrease in the over-all rate. The rate of responding began to increase in the next session, however, reaching the final performance shown in Fig. 27, 13 sessions later. The rate of responding was low at the start of the session; but it increased briefly early in the session in the region *a* to *d*, with responding during the time outs at *b* and *c*. For the remainder of the session the rate fell, reaching a value of approximately that obtained earlier in Fig. 26H.

Number 67. When No. 67's responses spaced less than 4 seconds apart were punished (Record J of Fig. 23), the first

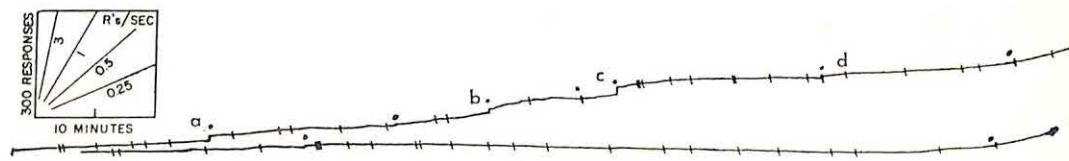


FIG. 27. Chimpanzee No. 97. Responses spaced less than 4 seconds are punished. Final performance.

effect was a decline in the rate of responding, followed by an increase in the rate five sessions later, in Record K. Note, however, that the rate of responding was already considerably below the critical value at which responses would be punished, and the increase in rate of responding occurring in Record K resulted in only two time outs for the entire session.

Interresponse-Time Records

Number 97. The punishment of responses spaced less than 1 second during the 20th to 26th sessions reduced the frequency of responses spaced less than 1 second to zero by the 26th session (Fig. 25). Responses spaced less than 2 seconds also declined markedly, but those spaced less than 4 seconds declined less. The remaining IRT classes also showed a decline in frequency, but only of the order of 5%, and reflected an over-all fall in rate rather than selective action of the punishment.

The punishment of responses spaced 2 seconds or less (Sessions 27-39) had little effect on the frequency of 2-second interresponse times, since this frequency was already near zero. As in the 1-second case, the effect of the punishment was selective, with IRT's nearest the punished one declining more than the longer IRT's. By the 39th session, over 30% of the responses were spaced *more* than 18 seconds from a previous response.

The punishment of responses spaced 4 seconds or less, beginning in Session 41, produced a further decline in the rate of responding, reaching the lowest value in Session 48. Almost 95% of the responses emitted in the session were spaced *more* than 18 seconds, even though only those responses spaced 4 seconds or less were punished. The decline in rate of re-

sponding occurred approximately uniformly in all the IRT classes. The increase in rate with further exposure to the punishment occurred in Sessions 48 to 54. The frequency of responses in the punished category ranged from zero to 0.04% and showed the least change. Responses spaced between 8 and 10 seconds showed the greatest change.

Number 67. In the case of No. 67 (Fig. 24) the punishment of responses spaced less than 1 second had little effect on these responses since the frequency was already near zero. The over-all rate of responding fell sharply, however, with no differential effect in any of the IRT categories under 18 seconds. The rapid decline in rate of responding continued as the criterion for punishment was increased to responses spaced 2 seconds; it reached the lowest value on the 29th session. Over 60% of the responses in this session were spaced more than 18 seconds, and 35% were spaced from 8 to 18 seconds.

The over-all rate of responding increased with further exposure to the punishment of 2-second IRT's and was reflected about equally in all of the IRT categories between 8 and 18 seconds. The increase in rate was not sustained, however, and the rate fell to a value somewhat higher than that prevailing previously in the vicinity of the 29th session.

The punishment of responses spaced 4 seconds or less, beginning on the 41st session, was followed by a temporary increase in the frequency of responses spaced 4 seconds or less, as well as a decrease in the frequency of responses spaced more than 14 seconds. With further exposure, responses spaced less than 10 seconds remained the same, along with an increase in the frequency of responses spaced *more* than 18 seconds.

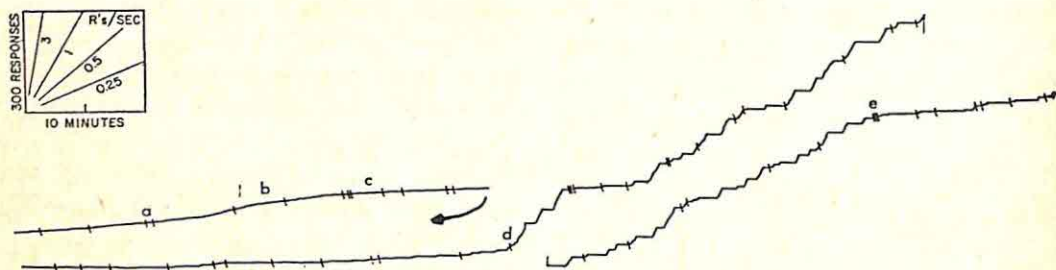


FIG. 28. Chimpanzee No. 97. Removal of the punishment. First session following Fig. 27.

Removal and Reinstatement of the Punishment Contingency

Number 97. On the session following the performance in Fig. 26, the punishment was discontinued and the procedure remained simply variable-interval reinforcement. Figure 28 shows the entire first session. The session began with the characteristic low rate until the region between *a* and *b*, when a slight increase occurred followed by a return to the prevailing low rate. After approximately 80 minutes, the over-all rate suddenly increased in the region *d* to a value almost comparable with that prevailing on this variable-interval schedule before the introduction of the punishment procedure. Local rates of responding reached 1.3 to 1.8 responses per second, and the grain of the record consisted of sustained bursts of responding separated by abrupt shifts to zero rates. Before the end of the session, however, beginning at *e*, the rate

fell again. Following the performance shown in Fig. 28, the over-all rate of responding fell markedly, but subsequently increased, developing into a performance shown in Record A of Fig. 29, taken after 15 sessions without punishment. The highest over-all rate of responding was slightly lower than that which prevailed at the start of the session before the punishment procedure. The total amount of responding increased, however, because the high initial rate was sustained longer before it fell, as, for example, in the region *e*. The fine grain of the record shows frequent occurrences of smooth positively accelerated segments, terminating abruptly in a pause of several minutes, as at *a*, *b*, *c*, *d*, and *f*.

The interresponse-time analysis (see Fig. 25) shows residual effects of the previous punishment. The frequency of responses spaced less than 0.33 second

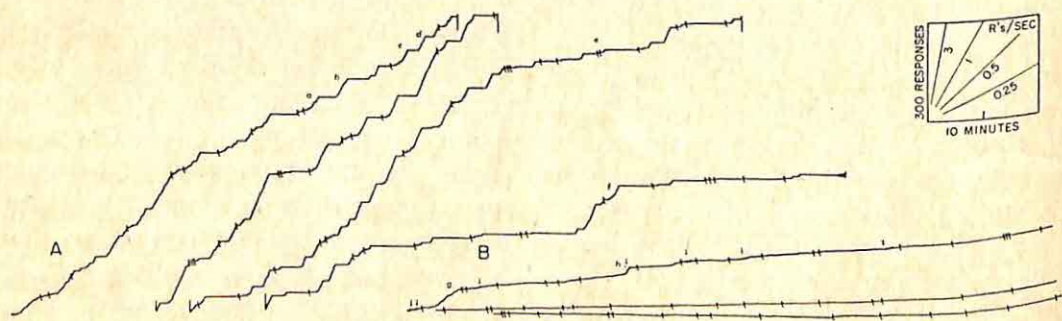


FIG. 29. Chimpanzee No. 97. Final performance without punishment (Record A) and the return to the punishment of responses spaced less than 4 seconds (Record B).

remained near zero, and responses spaced less than 1 second were less frequent than before the punishment. In the upper range of IRT's a substantial number of responses spaced more than 18 seconds occurred, compared with the number in the first nine sessions of the experiment. Most of the responses were spaced less than 2 seconds, or more than 12 seconds, compared with a more even distribution at the start of the experiment. The experiment was interrupted for five weeks between Sessions 58 and 59 with little effect.

On the session following Fig. 29, the punishment procedure was reinstated. The result is shown in its entirety in Record B. Responses spaced less than 4 seconds were again followed by a time out. The rate of responding was low for the first few seconds of the session, similar to the start of the previous session (Record A). Time outs occurred during this period; and, after a brief increase in the rate for approximately 100 responses, the rate of responding fell to a low enough value for the rest of the session so that only six time outs occurred during the first part of the session. The rate became even lower for the latter part of the session, with no responding occurring rapidly enough to qualify for punishment. The rate of responding was high enough, however, so that reinforcements were still being received on schedule.

Number 67. The result of removing and reinstating the punishment procedure was similar for No. 97, except for a slower increase in the over-all rate of responding when the punishment of fast responding was first removed. The cumulative curves of Fig. 30 describe the transitions to the variable-interval performances without time out and the subsequent return to punishment of responses spaced 4 seconds. Record A contains the

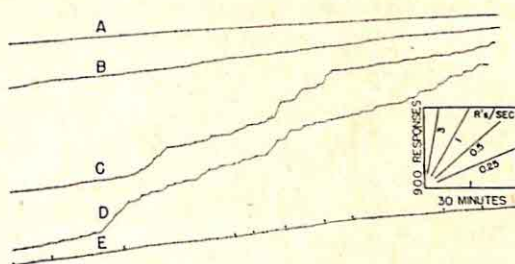


FIG. 30. Chimpanzee No. 67. Removal of the punishment and the subsequent reinstatement. A, B, C, and D. Ninth, tenth, fifteenth, and nineteenth sessions without punishment. E. Reinstatement of the punishment of responses spaced less than 4 seconds.

ninth session after the removal of the punishment. The over-all rate of responding was essentially that prevailing under the previous punishment procedure, in spite of an increase in the frequency of responses spaced zero to 2 seconds.

The rate of responding increased somewhat in the next session, shown in Record B; and by the 15th session without punishment the over-all rate increased more substantially (Record C), reaching almost the final performance shown in Record D on the 19th session. The reinstatement of the punishment contingency at the start of Record E produced an immediate fall in rate to a value approximately the same as it was previously (see Record K, Fig. 23).

Discussion

Results of the experiment show that the selective punishment of fast responding will reduce the over-all rate of responding below that normally generated by the variable-interval schedule, and that the effect is reversible, with the approximate return of the original schedule performance when responses are no longer punished. The effect of the punishment is not entirely specific to those IRT's that are punished. The punish-

ment of a given rate reduces the rate of responding considerably beyond the value at which no responses would be punished, suggesting a spread of effect of the punishment similar to the effects of electric shock. This is illustrated by the initial effect of the 4-second punishment on No. 97, where the rate of responding fell considerably below a response every 4 seconds, so that the subsequent rate increase had little effect on the number of responses punished.

The effect of lowering the rate of responding by punishment of fast responses contrasts with the procedure of lowering the over-all rate of responding by withholding reinforcement until a low enough rate occurs (differential reinforcement of low rate); the over-all rate of responding remains slightly higher than the required rate, with the differential rate requirement being met by frequent instances of spaced responses. However, the tendency for the effect of the time out to generalize might disappear with longer exposure to a given value. This is suggested by the tendency for the punishment to initially depress the rate of responding below the value necessary to avoid punishment. The subsequent increase in rate occurred without any significant increase in the number of punished responses.

Limiting values probably exist beyond which the punishment contingency would be ineffective in reducing the rate of responding. The maximum value is probably a function of the length of the time out and the level of deprivation. Another limitation might be the ability of the animal to react differentially to the passage of time. In differential reinforcement of low rates by withholding reinforcement, for example, the animal's performance approximates the contingency better at higher rates. The per-

formance of No. 67, during the punishment of responses spaced 4 seconds or less, suggests that 4 seconds is a limiting value for this animal under these conditions.

A low rate of responding was generated in this experiment by a progressive lowering of the rate. The contingencies in the situation suggest that a progressive program is most efficient; but it is not certain that initial prolonged exposure to an extreme value would not also be effective.

The large order of control by punishment with the time out from the same schedule of reinforcement used in Exp. I, II, and III gives further support to the findings of these experiments that the time out has two effects. The high rate in the pre-time-out stimulus occurs in spite of the general aversive properties of the time out demonstrated here and in Exp. IV.

Experiments with pigeons, using similar procedures, failed to confirm this result. The failure of the time out to suppress fast responding in the pigeon does not imply absence of aversive properties of the time out. In the face of other kinds of aversive control which have been established (Exp. III), the time out will probably suppress fast responding if the proper conditions are found. The difficulty is probably related to the effect of the time out in increasing the rate of responding, and the prepotency of this effect of the time out over the aversive properties.

SUMMARY

Experiment I. A 30-second pre-time-out stimulus periodically preceded a 10-minute time out from a 3-minute variable-interval schedule of reinforcement. The rate of responding during the pre-time-out stimulus increased relative to

the over-all rate of responding. This increased rate was interpreted as a result of differential rate reinforcement by the selective effect of the time out on the spacing of reinforced responses.

Experiment II. The same variable-interval schedule of Exp. I was interrupted every minute by a 3-minute time out. The result was a large-order-of-magnitude rate increase compared with the uninterrupted performance. This function of the time out was analyzed in terms of its relation to the effect of the time out in Exp. I.

Experiment III. The aversive properties of the time out of Exp. I and II were emphasized by a procedure permitting the pre-time-out stimulus to terminate without a time out if the rate of responding were low enough toward the end of the pre-time-out stimulus. Under this procedure the rate during the pre-time-out stimulus fell to zero. Subsequent exposure to the same procedure produced a performance in which the rate of responding in the pre-time-out stimulus was negatively accelerated. Time outs did not occur because the rate of re-

sponding reached zero by the end of the pre-time-out stimulus.

Experiment IV. The generality of the aversive properties of the time out from the variable-interval schedule of Exp. I, II, and III was extended by generating avoidance behavior based on the postponement of the time out; the variable-interval schedule of food reinforcement of responses on one key was interrupted periodically by a time out unless a second key was pressed. Each press of the second key postponed the time out. Substantial levels of avoidance behavior were recorded at several values of length of postponement of the time out.

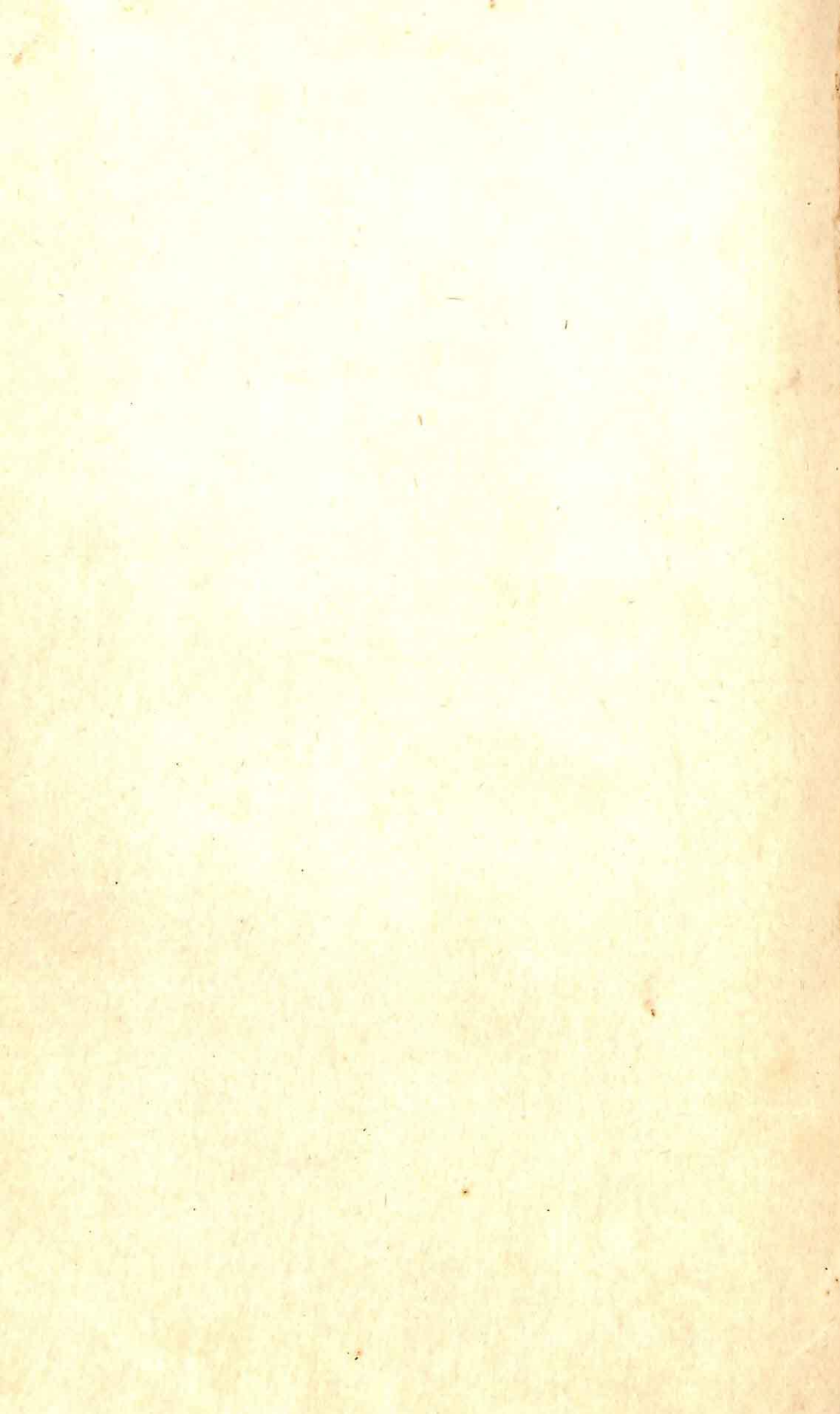
Experiment V. An examination of the properties of the time out used as punishment further established the generality of the aversive properties of the time out from the variable-interval schedule of Exp. I, II, III, and IV. The rate of responding on the variable-interval schedule was reduced when a time out selectively followed "fast responding." Responses of various interresponse times were punished with a resulting change in the rate of responding.

REFERENCES

1. AZRIN, N. Some effects of two intermittent schedules of immediate and non-immediate punishment. *J. Psychol.*, 1956, **42**, 3-21.
2. BRADY, J. V., & HUNT, H. F. An experimental approach to the analysis of emotional behavior. *J. Psychol.*, 1955, **40**, 313-324.
3. DEWS, P. B. Studies on behavior. I. Differential sensitivity to pentobarbital of pecking performance in pigeons depending on the schedule of reward. *J. Pharmacol.*, 1955, **113**, 9.
4. DINSMOOR, J. A. Punishment: I. The avoidance hypothesis. *Psychol. Rev.*, 1954, **61**, 34-46.
5. ESTES, W. K., & SKINNER, B. F. Some quantitative properties of anxiety. *J. exp. Psychol.*, 1941, **29**, 390-400.
6. FERSTER, C. B. The use of the free operant in the analysis of behavior. *Psychol. Bull.*, 1953, **50**, 263-274.
7. FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement*. New York: Appleton-Century-Crofts, 1957.
8. HERRNSTEIN, R. J. Behavioral consequence of removal of a discriminative stimulus associated with variable-interval reinforcement. Unpublished doctoral thesis, Harvard Univ., 1955.
9. HERRNSTEIN, R. J., & MORSE, W. H. Escape behavior maintained by the removal of a stimulus correlated with an intermittent schedule of food reinforcement. Paper read at Eastern Psychol. Assn., Atlantic City, March, 1956.
10. HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
11. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

12. MORSE, W. H. An analysis of responding in the presence of a stimulus correlated with periods of nonreinforcement. Unpublished doctoral thesis, Harvard Univer., 1955.
13. MORSE, W. H., & HERRNSTEIN, R. J. The maintenance of avoidance behavior using the removal of a conditioned positive reinforcer as the aversive stimulus. *Amer. Psychologist*, 1956, 11, 430. (Abstract)
14. SCHOENFELD, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. H. Hoch & J. Zubin (Eds.), *Anxiety*. New York: Grune and Stratton, 1950. Pp. 70-99.
15. SIDMAN, M. Two temporal parameters of the maintenance of avoidance behavior by the white rat. *J. comp. physiol. Psychol.*, 1953, 46, 253-261.
16. SIDMAN, M. Technique for assessing the effects of drugs on timing behavior. *Science*, 1955, 122, 925.
17. SKINNER, B. F. Superstition in the pigeon. *J. exp. Psychol.*, 1948, 38, 168-172.
18. SKINNER, B. F. How to teach animals. *Sci. Amer.*, 1951, 185(6), 26-29.
19. WILSON, M. P., & KELLER, F. S. On the selective reinforcement of spaced responding. *J. comp. physiol. Psychol.*, 1953, 46, 190-193.

(Accepted for publication March 18, 1957)



Rigidity-Flexibility and Intelligence: A Cross-Sectional Study of the Adult Life Span from 20 to 70 Years¹

K. WARNER SCHAIE
University of Washington²

RECENT thinking in the biological sciences, based on extensive case analyses in clinical medicine as well as searching actuarial studies of nationwide morbidity patterns, seems to indicate important relationships between certain behavior patterns and the decline in functional efficiency and structural maintenance of the human organism. From such a biological view, a trend seems to emerge which would view the normal aging process as a decrease in the excitability of tissue to stimuli, the organism being increasingly less able to adjust to the changes in reaction potential required by a given life situation (17, 28).

While a number of interesting physiological hypotheses for such a view might

be considered at the molecular level, there is greater current concern at the molar level with types of explanation which require information on a multitude of purely psychological processes. In terms of the prediction of the rate and pattern of such molar change, psychological tools thus assume primary importance.

Until comparatively recent years, psychologists working in the developmental area have been primarily concerned with the periods of childhood and adolescence. Behavior theorists give considerable emphasis to a description of childhood behavior and of adolescent conflict, but take little or no cognizance of the conflicts, crises, and behaviorial changes of the adult life span. Excepting perhaps Erikson (6), no formal or theoretical discussion is to be found dealing with the phenomena of behavior change in the adult from the period of full maturity to senescent decline, even though descriptive and anecdotal material abounds.

In the present study a conceptual model for changes in adult behavior will be considered. Since this model is empirically derived, certain methodologies for its verification will be discussed and a major experiment dealing with age changes in intelligence, which bears on this model, will be described.

To begin with, it may be suggested that one way in which the human life cycle could be analyzed is to utilize the

¹ This monograph is based upon a doctoral dissertation at the University of Washington, Seattle (22). I owe much gratitude to Charles R. Strother and Paul Horst for their advice throughout the progress of this study. Acknowledgments need also be made to Read Tuddenham, Fred Rosenthal, and Ralph T. Granneberg who aided in various ways when the problem was first conceived. H. G. Gough permitted the use of his Rigidity Scale as part of our test battery with the cooperation of Consulting Psychologists Press, Palo Alto. The staff of Group Health Cooperative of Puget Sound were most cooperative in making available their records and assistance in drawing the sample of subjects. Specific acknowledgment should go to E. R. Clark, John Switzer, and Cecil Arnold. Finally, I want to express my gratitude to my wife without whose help and patience this study would have been impossible.

² The author's present address is: Department of Psychology, University of Nebraska, Lincoln, Nebraska.

concept of rigid and flexible behavior. A recent paper by Doerken (4), for example, suggests that the child starts out from a rather constricted environment, that during the period of maturation a widening of the child's constricted world occurs, and that, as a result of the increased opportunity of choice and an increasingly permissive environment, the individual generally reaches maturity as a rather flexible individual. From then on, the individual's "lifespace" again tends to constrict progressively until in old age a situation somewhat similar to early childhood experience is reached (3). Some authors further contend that possession of superior mental abilities may slow down the process of constriction and maintain that the higher an individual's intelligence the longer will he be able to function in a relatively flexible and functionally efficient manner (7).

If the above formulations are correct, one might be able to derive prediction equations from an empirical assessment of the rate of change in both intelligence and rigidity which could be used to some extent to make long-range predictions about the expected decline of an individual's functioning.

Proceeding a step further, one might examine some behavioral concomitants of test patterns indicative of flexible or rigid behavior at given age intervals. Such a procedure might suggest cut-off points for the prediction of functional efficiency and also aid the counselor in suggesting to his client the kind of activities that seem to be associated with flexible behavior and which ought to be encouraged to slow the constrictive process.

The present study is the culmination of a lengthy experimental program. In view of the dearth of material on be-

havioral change over the adult life span, it seemed necessary first to test and develop instruments which would be useful over a wide range of response both in terms of ability and age suitability. The first study in this program was conducted in 1951 (together with F. Rosenthal and R. M. Perlman) and was concerned with testing the suitability and reliability of the intermediate form of Thurstone's SRA Primary Mental Abilities test for older adults (26). The second more extensive study was concerned with the construction of an equally suitable battery for the measurement of behavioral rigidity and the assessment and validation of its factorial structure (21). With assurance in the usefulness of these devices, it was possible to proceed with the further analysis of adult behavior change.

A CONCEPTUAL MODEL FOR THE ANALYSIS OF AGE CHANGES IN ADULT BEHAVIOR IN TERMS OF RIGIDITY-FLEXIBILITY

A hypothetical definition of aging has previously been suggested in terms of increasing resistance of the organism to behavioral or environmental change. This rather loose statement needs to be enlarged upon before one can deduce behaviorally verifiable consequences.

If one considers some of the conventional age gradients for many traits and abilities, one finds a fairly rapid rate of change during the early maturational period and a gradual slowing down as the developmental peak is reached and passed. From then on the developmental process is characterized by a gradual decrease in the effectiveness and personal adjustment of most individuals. Enough individual variation is found at each age level, however, to demand the introduction of additional explanatory factors.

If socially effective and mature be-

havior is taken to be a function of the individual's ability to adjust without difficulty to changes in the pattern of cognitive and social stimuli which make up his environment, one might then consider these individual differences to be an effect of a more or less complex interaction of initial potential ability and changing ability to utilize this potential for the acquisition of new behavior patterns.

If this formulation is correct, one might then go further to suggest that the individual who is most able to adjust to the stress imposed by changes in the patterns of environmental stimuli will tend to respond in a socially effective manner longer than his less flexible peer. Given reasonably constant environmental conditions it might be possible to measure a person's ability to adjust to change, once he reaches maturity, and to predict his personal gradient of decline in effective behavior.

The important question arises, however, as to what kind of maturational gradient this presumably central factor of ability to adjust to change may have. The impact of the postulated central factor would obviously depend upon the form which its age gradient might take. Alternate hypotheses will be proposed to explain age changes in behavior which are suggested by some of the most likely forms of such a gradient, but first it will be useful to consider how the ability to adjust to change might best be defined for purposes of measurement.

The class of behavior which has been closely studied in the past and which approximates fairly well the psychological aspects of adjustment to change has usually been subsumed under the heading of "behavioral rigidity." While there are almost no studies of a developmental nature dealing with rigidity, a great

amount of work has been done to investigate the dimensions of this concept as well as to demonstrate their relation to specific personality configurations and to patterns of normal and abnormal behavior.³

In an earlier paper, it was found necessary to formulate a summary definition for the many concepts advanced as criteria of such behavioral rigidity. Behavioral rigidity was there defined as:

... a tendency to persevere and resist conceptual change, to resist the acquisition of new patterns of behavior, and to refuse to relinquish old and established behavior patterns (21).

This definition may seem rather broad, but a set of behavioral consequences can readily be deduced which are amenable to observational test.

From such a definition one can deduce, for example, that a rigid person dislikes to change patterns of interpersonal relationships, has difficulty in changing from one activity to another, is less efficient in activities requiring alternation and continuous readjustment of motor activities, etc.

It was shown elsewhere (19) that a battery of tests could be constructed to cover a great variety of operations subsumed under the above definition of behavioral rigidity. As the result of a series of factor analyses (21) it was further possible to show that this concept could best be described by a three-dimensional system. In terms of the tests on which high loadings for the three factors in this system were found, these dimensions were named "motor-cognitive rigidity," "personality-perceptual rigidity," and "psychomotor speed." An inspection of the nature of the tests high on these factors indicate

³ Surveys of the literature on different aspects of the concept of "rigidity" may be found in papers by Werner (33) and by Levine (15); see also (22).

that they just about cover the gamut of psychological views on behavioral rigidity.

In terms of the operations specified by the tests identifying these factors the following definitions seem appropriate: The motor-cognitive rigidity factor seems to indicate the individual's ability to shift without difficulty from one activity to another; it is a measure of effective adjustment to shifts in *familiar* patterns and to continuously changing situational demands. The personality-perceptual rigidity factor seems to indicate the individual's ability to adjust readily to *new* surroundings and change in cognitive and environmental patterns; it seems to be a measure of ability to perceive and adjust to *unfamiliar* and *new* patterns and situations. The psychomotor speed factor finally indicates an individual's rate of emission of familiar cognitive responses. A high score on this factor would seem to imply superior functioning efficiency in coping with familiar situations requiring rapid response and quick thinking.

As a result of these studies some other "hunches" emerged which led to further preliminary investigations. For example, when the records of a sample of 216 subjects were rescored in terms of these derived factors, it was possible to show that rigidity in terms of all three factors increases with age in a manner similar to most conventional age gradients. An analysis of variance of two groups of 20 subjects, one aged below 25 and the other above 55, showed significant differences on all three factors at the 1% level of confidence. Furthermore, there appeared to be a significant correlation with an "intelligence" factor also extracted in the factor analysis (19). Great variation was found, however, among individuals, so that many old individuals were found both "intelligent" and "flexible," while

some young individuals were found who were rigid.

This line of inquiry suggests that the level of functioning of an individual with respect to cognitive and perhaps other behavior traits may be predictable in terms of a more central factor which might be measured by a three-dimensional rigidity factor system and possibly well approximated by a composite score formed by a linear combination of the three factor scores.

Such a formulation might also be useful for the prediction of adult age changes in various types of socially important behavior classes, provided it can be shown that not only level of functioning, but maintenance of basal level of functioning over the adult age span can be meaningfully related to this central factor.

It was therefore felt desirable to draw a reasonably large sample, stratified cross-sectionally across the major portion of the adult age range, to assess changes in the degree of behavioral rigidity noted for successive age samples, and, further, to inquire what the relation of this variable might be with other important variables at various age intervals.

Since the maintenance of intellectual ability is of such paramount importance in socially effective behavior, and because of the positive relationship found in the preliminary studies, it was decided to concentrate upon an analysis of the relationship between behavioral rigidity and Thurstone's Primary Mental Abilities throughout the major portion of the adult age span. Such an inquiry has both practical and theoretical implications. For both, however, one must study the form of the age gradients as well as their interrelation to gain more insight as to where the crucial points of behavioral change might lie.

It has been stated earlier that the pre-

dictions which can be made about behavior change from a knowledge of performance on the rigidity variables will depend largely upon the form of the maturational changes shown by these variables. Depending on the form of this gradient it will be possible to examine four alternatives, which are listed here in descending order of predictive potency:

I. Behavioral age changes are a function of the individual's lack of ability to adjust to change and can thus be predicted for an advanced age from a knowledge of his standing on the rigidity-flexibility dimension at the basal age.

II. Behavioral age changes occur concomitantly with decreasing adjustment to change, and the level of function for individuals at any given age level is therefore predictable from knowledge of the rigidity-flexibility dimension.

III. Behavioral age changes and changes in rigidity-flexibility show no significant relation whatsoever.

IV. A further alternative must be considered when Proposition I is falsified, i.e., lack of ability of an individual to adjust to change may be a function of some other behavior variable and may be predictable for his old age from a knowledge of that variable at the basal age.

An empirical description of the rate of change observed on the variables to be measured is of interest in any event; although, in terms of prediction, the second alternative would be less powerful, the third trivial, and the fourth would be disproof of our basic assumptions.

In line with the previous discussion it is timely now to state more precisely the hypotheses concerning the influence of behavioral rigidity upon the specific case of changes in the mental abilities with age:

Case I. If there is a random distribution on the rigidity-flexibility dimension for each level of ability at the age of developmental peak and if age changes in ability are a function of the scores on the rigidity-flexibility dimension, then one should find zero or modest correlation between rigidity and intelligence at the basal age with a systematic increase of correlation for successive age groups. There will be a family of gradients of decline for each level of ability (the steepest gradient prevailing for the most

rigid members) resulting in the mean values for the flexible members of a group remaining relatively constant while those for the rigid members would drop. In symbolic form:

$$BYf = BYr$$

$$\text{but } BOf > BOR$$

where BYf is the mean of ability measures for the young flexible members of the sample, BYr is the mean of ability measures for the young rigid members of the sample, BOf is the mean for the old flexible members, and BOR the mean for the old rigid members of the sample. The analysis of covariance may be used to test formally the validity of the first hypothesis.

Case II. If age changes in the mental abilities occur concomitantly with those in rigidity-flexibility, then either will serve to predict to some extent the level of functioning of the other at any given age level; and one should find significant but constant correlation throughout the adult age range influenced only by sampling variation and perhaps showing a slight drop towards the older end due to the effect of attrition. In this case the mean values for both the flexible and rigid members of the sample will show a consistent drop over the age range, but on the measures of ability there will be a significant but not increasing difference between the rigid and flexible members of the group at each age level, with the flexible members scoring in the positive direction. Again in symbolic form:

$$BYf > BYr$$

$$\text{and } BOf > BOR$$

The analysis of variance and covariance will be the appropriate test also for the second hypothesis.

Case III. If there are concomitant age changes in both rigidity-flexibility and the mental abilities but no significant relation exists, then the analysis of variance should verify the null hypothesis that rigid and flexible individuals do not differ in level of ability and the analysis of covariance should show that there is no relationship between differences on the rigidity-flexibility dimension and the age changes in mental abilities. In symbolic form it may be stated then that for this alternative:

$$BYf = BYr$$

$$\text{and } BOf = BOR$$

Case IV. Since this is the falsification of Case I, it requires identical conditions, viz.: If there is a random distribution on each mental ability for each level of rigidity-flexibility at the basal age and if age changes in rigidity-flexibility are a function of the scores on the Primary Mental

Abilities, then one should find zero or low correlations between rigidity and intelligence at the basal age with systematic increase of correlation for successive age groups. The mean rigidity-flexibility values for the highly able will remain relatively constant while the scores for those of low ability will drop. In symbolic form:

$$\begin{aligned} RYh &= RYl \\ \text{but } ROh &> ROl \end{aligned}$$

where RYh is the mean of rigidity-flexibility measures for the young highly intelligent members of the sample, RYl the mean for the young people of low intelligence, ROh for the old highly intelligent, and ROl for the oldsters of low intelligence. The analysis of covariance will again be the appropriate test.

Further discussion of the implications of these hypotheses is reserved until after the empirical evidence is examined. The method and sample of subjects used to investigate these propositions will next be described.

THE TEST MATERIAL

The Assessment of Behavioral Rigidity

As already mentioned, it was decided to utilize the results of an earlier inquiry (19, 21), concerned with determining the dimensions of behavioral rigidity, in providing the measures of rigidity for the present study. In these earlier studies a battery of tests of rigidity, some of which were selected from the literature and others of which were newly constructed, had been factor analyzed. Since developmental studies were anticipated, the final test battery retained only those tests which appeared applicable to young adults as well as mature and senescent populations. For practical reasons, the battery was further restricted to tests suitable for group administration and an effort was made to retain those tests where social status and educational level were of minimal importance in influencing test results.

The factor analysis of these tests, using a modified version of the multiple-group method

(12), resulted in the isolation of three factors termed motor-cognitive rigidity, personality-perceptual rigidity, and psychomotor speed. The factorial structure of the test battery was further cross-validated by drawing a new sample and factoring the correlation matrix for this sample. The oblique factor loadings for both analyses (reported in 21) were very similar and indicate rather stable structure for the hypothesized dimensions.

The original test battery was shortened by deleting the two tests which contributed the lowest factor loadings to their respective factor scores. The remaining four subtests yield seven scores and require about 30 minutes for administration. The following paragraphs give a brief description of the Test of Behavioral Rigidity (hereafter called TBR).⁴

1. The Capitals Test. The S copies a paragraph first as it is printed on a sample card and then once more, substituting capitals for lower case letters and vice versa; e.g. Original: "The Duke DREW his sword." Second time: "tHE dUKE drew HIS SWORD." Two and one-half minutes are allowed for each part, and an estimate of rigidity is obtained by dividing the number of correctly written words in the test series by the number of correctly written words in the practice series. The number of words in the practice series is used as one of the psychomotor speed estimates.

2. The Opposites Test. This test contains three series of 40 stimulus words each. In the first series, the S is required to respond by giving the opposite of the words listed. In the second list, he responds with a synonym of the word listed. The last list finally presents a mixture of stimuli to which S must respond either with the synonym or the antonym. When the stimulus word is printed in capitals, the synonym is required; while when small letters are used, the antonym is the required response. The number of correct responses on the first two series gives another psychomotor score. Two rigidity scores are yielded by this test. The first measures accurate cognitive shift and is based on the third series only. It is given by dividing the number of rigid (i.e. inappropriate) responses by the total number of responses given on this series. The second rigidity score measures the influence of rigidity in inhibiting rapid cognitive shift and is given by the formula: number of correct re-

⁴ Copies of the test booklet and the manual of instructions for administration and scoring have been deposited with American Documentation Institute, Order Document No. 5573, remitting \$2.00 for 35-mm microfilm or \$3.75 for 6-by-8 in. photocopies.

sponses in the third series divided by one-half the number of correct responses in Series 1 and 2. Two minutes are allowed for each series.

3. Scale of Rigidity (R scale). This is a 22-item scale of general statements which Gough (8) found to be empirically related to rigid behavior. These items are mixed among 44 masking items which in themselves can be scaled as a measure of social responsibility (10). The score is the number of flexible responses favorably endorsed by the S.

4. The Rigidity Index (P scale). This is a 9-item inquiry as to rigid modes of behavior. It contains questions like: "Do you feel strongly inclined to finish whatever you are doing in spite of being tired of doing it?" The score is also the number of flexible responses. (In the test booklet the P scale appears as items 151-159 of the Scale of Rigidity.)

The Capitals Test was adopted from Bernstein's extensive study of quickness and intelligence (2) and is representative of the "functional" or Spearmanian approach to the study of perseveration and rigidity. The Opposites Test was newly constructed by the writer, along lines suggested by the work of Scheier and Ferguson (27). All the items of the R scale were obtained from Gough's California Psychological Inventory (9), although some of the statements and the arrangement of items were modified for use throughout the adult age range by Schaie and Rosenthal (25). The Rigidity Index, finally, was adapted by the writer by means of Guttman scale analysis technique (11) from a 17-item scale first used by Lankes (14).

The present battery meets one of the most common criticisms of rigidity tests (16, 32) by recording two separate scores in the tests where speed is at issue. The first score on the Capitals and the Opposites Tests (termed the "non-rigidity"—NR—score) represents performance on a practice series and turns out to be a good estimate of the psychomotor speed dimension. The second score is the "rigidity" score, i.e. the score indicating the amount of interference resulting when rigid behavior inhibits adjustment to the altered condition.

In computing the rigidity scores on the tests where time limits were imposed, a correction formula was applied to allow for irrelevancies introduced into some of the tests due to motor speed and difficulty of alternation. Since the influence of specific test function has been allowed for by the two separate scores, these artifacts may be corrected simply by scoring the ratio between the number of responses on the practice and the test series. To permit uniform interpretation this scoring system is so arranged as to assign high scores for flexible and low scores for rigid performance.

TABLE 1
WEIGHTS USED FOR TRANSFORMING DEVIATION
SCORES ON THE TBR SUBTESTS TO
FACTOR SCORES

	Motor- cogni- tive rigidity	Person- ality- percep- tual rigidity	Psycho- motor speed	Com- posite rigidity
1. Capitals- R	.25			.083
2. Capitals- NR			.60	.200
3. Opposites 1	.35			.117
4. Opposites 2	.40			.133
5. Opposites- NR			.40	.133
6. R scale		.50		.1665
7. P scale		.50		.1665

Once the factorial structure of the test battery was known to be essentially three dimensional, it was possible to write appropriate regression equations to permit an estimate of factor scores from knowledge of the test scores of subjects on the original variables having high loadings on the specified factors. The deviation scores for the subtests may then be appropriately weighted and combined to yield factor scores. Since two estimates of factor loadings were available (see 21), a rounded-off average was taken to provide the regression weights for the purpose of constructing transformation tables for the revised test battery. These weights are given in Table 1.

Since rigidity scores were to be assigned such that they would be easily comparable to other tests, it seemed convenient to write them in the form of standard scores with a Mean of 50 and an SD of 10 (a further transformation to a "rigidity quotient" form with a Mean of 100 and an SD of 15 is given in the manual of instruction).⁴

The weights for the composite rigidity score represent a simple additive linear combination. This weighting was used for computational simplicity since a least-squares approximation was shown to be only negligibly better when multiple correlations between the three factor scores and a composite IQ score were computed.

While norms were available for various population samples (20), it was decided to apply the above mentioned transformation to the deviation scores computed on the basis of the present study, since it provides a larger and more representative sample than any of the previously used populations.

The Assessment of Mental Abilities

One of the common criticisms of tests of intelligence has been the lack of readily identifiable subtest meaning and the high subtest intercorrelations. It was felt that while the Wechsler-Bellevue, for example, may be a useful measuring device for clinical purposes, there are other research tests of more clear-cut factorial structure which, in addition to a global intelligence score, would also yield relatively independent estimates of distinct mental abilities.

Thurstone's Primary Mental Abilities Test, Intermediate Form (hereafter called PMA) (30) seemed the best instrument for the purposes of this study. Extensive normative research has been done on this test with adolescents and previous work has indicated its applicability to a sample of Ss aged 53-78 years (26). The test contains relatively little material related to formal education and thus seemed well suited for this study. The PMA test battery was derived by factor analysis of about 60 mental tests, resulting in the isolation of several distinguishable mental abilities (29). A test was then constructed to measure each of these abilities and their relative independence was demonstrated in further investigations by low intercorrelations (13).

The PMA test battery consists of five subtests designed to cover the presumably independent factors listed below with the description furnished in the test manual (31).

1. Verbal-Meaning (V) is the ability to understand ideas expressed in words. It is used in any activities in which information is obtained by reading or listening to words.

2. Space (S) is the ability to think about objects in two or three dimensions. It is perhaps best described as the ability to imagine how an object or figure would look when it is rotated, to visualize objects in two or three dimensions, and to see the relations of an arrangement of objects in space.

3. Reasoning (R) is the ability to solve logical problems—to foresee and plan. It is probably the most important of the mental abilities. The person with good reasoning ability can solve problems, foresee consequences, analyze a situation on the basis of past experience, and can make and carry out plans according to recognizable facts.

4. Number (N) is the ability to work with figures—to handle simple quantitative problems

rapidly and accurately. It involves primarily speed and accuracy in handling numbers.

5. Word-Fluency (W) is the ability to write and talk easily. It differs from Verbal-Meaning because it concerns the speed and ease with which words can be used, rather than with the degree of understanding of verbal concepts.

In addition to the factor scores it is possible to obtain a composite IQ equivalent score by summing the subtest scores according to the formula:

$$\text{Composite} = V + S + 2R + 2N + W$$

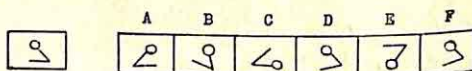
The differential weights, according to Thurstone, were assigned in inverse proportion to the standard deviation of the tests to give each test approximately equal importance.

The following paragraphs describe the material used in the PMA subtests:

VERBAL-MEANING is measured by a test made up of the following type of item:

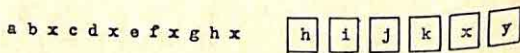
ANCIENT: A. Dry B. Long C. Happy D. Old. The word which means the same as the first word is to be marked. There are 5 practice items and 50 test items. Four minutes are allowed for this test.

SPACE is measured by 20 items similar to the one below:



Every figure in the row that is the same as the first figure, even though it is rotated, is to be marked. Figures that are mirror images of the first figure are not to be marked. There are six practice items. Time for this test is five minutes.

REASONING is measured by items like the following:



The letters in the row form a series based on a rule. The problem is to mark the letter that should come next in the series. The test is made up of 10 practice items and 30 test items, and the time limit is six minutes.

NUMBER is measured by a test with items of the following kind:

17		
84		
29		
140		

The sum of each column of figures is given, but some of the solutions are right and others

are wrong. Five practice items and 60 test items are given, with a time limit of six minutes.

WORD-FLUENCY is measured by a test requiring the writing of as many words beginning with a certain letter as possible in a specified time. The letters S or M are usually used and the time limit is five minutes.

Since the test was originally constructed for a high school age population, Schaie, Rosenthal, and Perlman (26) were concerned whether the high reliability and low intercorrelation data claimed for the test by its authors would hold up for an adult sample. Their results indicate that the factors presumably measured by the tests remained equally independent for an old adult group. They also obtained reliability data by administering the test as a power test and obtained reliability coefficients ranging from .94 to .96, indicating that the PMA is a stable test also for old adults.

Norms for the PMA are available only for a high school population. It was therefore decided to transform all obtained scores as well as the composite IQ score into standard scores with Means of 50 and SDs of 10, based upon the observed distribution of our own sample.

All test material was administered in a group testing situation. Twenty-four testing sessions were held in classrooms at the University of Washington and in various public schools throughout the King County area. Attendance at these sessions varied from about 15 to 45 examinees, with one examiner and one proctor. The testing sessions lasted approximately two hours and were handled in a standard manner.

SUBJECTS

Subjects for this study were obtained in the course of a cooperative research project from the membership of a pre-paid medical plan. The total pool of potential Ss over 16 years of age available from this organization consisted of about 18,000 members. A stratified random sampling procedure picked the names of 2,818 potential Ss who were then contacted and asked to volunteer. In response to this request positive replies were obtained from 451 men and 459 women, a total of 910 Ss.⁵ Due to con-

flicting time schedules and other sources of attrition, such as incomplete test records, the final sample consists of 500 Ss ranging in age from 20 to 70 years. The sample was divided into ten 5-year intervals, each containing 25 men and 25 women.

As seems inevitable in studies employing volunteer Ss, there was a tendency for the better educated levels of the pool of potential Ss to participate. In spite of this fact a great range of educational background was found. In the sample, 11.0% of the Ss had no more than a grade school education, 42.4% had attended or completed high school, 32.6% had attended or graduated from college, and 14.0% had done some postgraduate work in addition to their college education. The Mean number of years of education was 12.7 for the male and 13.0 for the female Ss. In terms of age differences, Mean number of years of schooling varied from 11.0 years for the 61-65 year-old group, which was lowest, to 14.1 years for the 26-30 year-old group. These differences are significant but would seem to reflect the increased educational opportunities available for the younger Ss, rather than differences in the population from which the Ss were obtained, since these changes run pretty close to the general census data. It should be noted, though, that the mean educational level for all subgroups is somewhat above that of the general population.

All Ss were rated as to their occupational level on a scale ranging from 0 to 9 using the major classifications of the United States Census Bureau, ranging from "unskilled labor" to "professional." Where the occupation "housewife" was given, the occupational rating of the husband was used. On this rating scale, the sample attained a Mean rating of 6.0 which is higher than the population average but is probably quite close to the population distribution in a metropolitan area like Seattle, where one would find a larger proportion of skilled mechanics, clerical workers, and professionals than would be true for the population average. According to this classification, the sample contains: 1.2% unskilled laborers, 0.8% employed in domestic service, 3.6% operatives, 5.0% employed in the service occupations, 5.8% in the protective services, 13.6% skilled craftsmen, 33.0% in sales or clerical work, 16.6% at the managerial or proprietary level, 15.2% engaged in semiprofessional and 5.2% in professional occupations. The Mean occupational levels for different age groups vary from 5.2 to 6.4 points. These age differences show a systematic trend, achieving a peak in the 46-

⁵ A more complete description of the known characteristics of the sample population as well as its relationship to the parent pool from which it was drawn will be published elsewhere.

50 year period and showing some decline thereafter.

All participants indicated their total family income on a rough 6-point scale. On this scale, 6.6% of the sample had a total family income below \$2,000, 18.8% were in the interval \$2,000 to \$4,000, 41.8% in the \$4,000 to \$6,000, 18.2% in the \$6,000 to \$8,000, 10.4% in the \$8,000 to \$10,000, and 4.2% in the above \$10,000 brackets. The Mean income level for the entire group was equivalent to a total annual family income of about \$4,500.

Since the source of Ss was a family medical plan, most of them are found to be married with a sprinkling of single, widowed, and divorced persons. In the sample, 6.6% of the group were single, 85.0% married, 2.6% divorced, and 5.8% had been widowed. The distribution of marital status is homogenous with respect to age, except for the youngest group (where there is a high incidence of people presumably before the stage of establishing families) and in the oldest groups (where evidence is found of family dissolution through the death of one of the partners).

From the above descriptive characteristics it would appear that the cross-sectional sample is reasonably representative of the broad middle group of our population. The sample is not quite as representative for the lowest and highest population segments, but in terms of the socioeconomic characteristics given is probably an adequate sample of the upper 75% of the American social structure. This is probably as good a population sample as can be constituted without the use of captive populations.⁵

RESULTS

Age Changes in Mental Abilities and in the Measures of Rigid Behavior

After all test records had been scored as prescribed by the examiner manuals (31, Footnote 4), the resulting raw scores were key punched into IBM cards. Means and standard deviations for the total sample, and separately for the ten 5-year age groups, were computed on the IBM Type 604 Calculating Punch according to methods described by Lunneborg,

Wright, and Ax (18). All raw scores were then converted into standard *T* scores with a Mean of 50 and an *SD* of 10, the statistics for the total sample being used as point of reference, which in view of the sample size would seem to give relatively stable estimates of the universe parameters. As a result, mean scores in the subsequent tabulations are comparable between different variables as well as between different age subsamples. New means and *SD*s were then computed for the subgroups and graphs were plotted to show the resulting age gradients.⁶

1. *Age changes in the measures of behavioral rigidity.* As was indicated in the section on the test material, the TBR version used in this study yields seven different subtest scores which, when combined according to appropriate weighting formulae, will yield three factor scores and a composite rigidity score which is a linear combination of these factor scores. Table 2 gives the means and *SD*s for the factor scores for each of the subgroups. Figure 1 gives the age gradients which result in graphic form.

Figure 1 and Table 2 indicate systematic age changes on all the rigidity dimensions. Analysis of variance tests show differences between age groups to be significant at the 1% level of confidence for every variable. It may thus be concluded on the basis of these data that people do become more rigid with increasing age on all the dimensions of rigidity measured.

To give a better view of the extent of change in these scores, the absolute decrement in *T* scores from the mean score of the most flexible group has been

⁶ In studying the following graphs and tables the reader might wish to keep in mind that differences between subgroups are significant if they exceed 2.63 *T*-score points at the 1% level and if they exceed 1.98 *T*-score points at the 5% level of confidence.

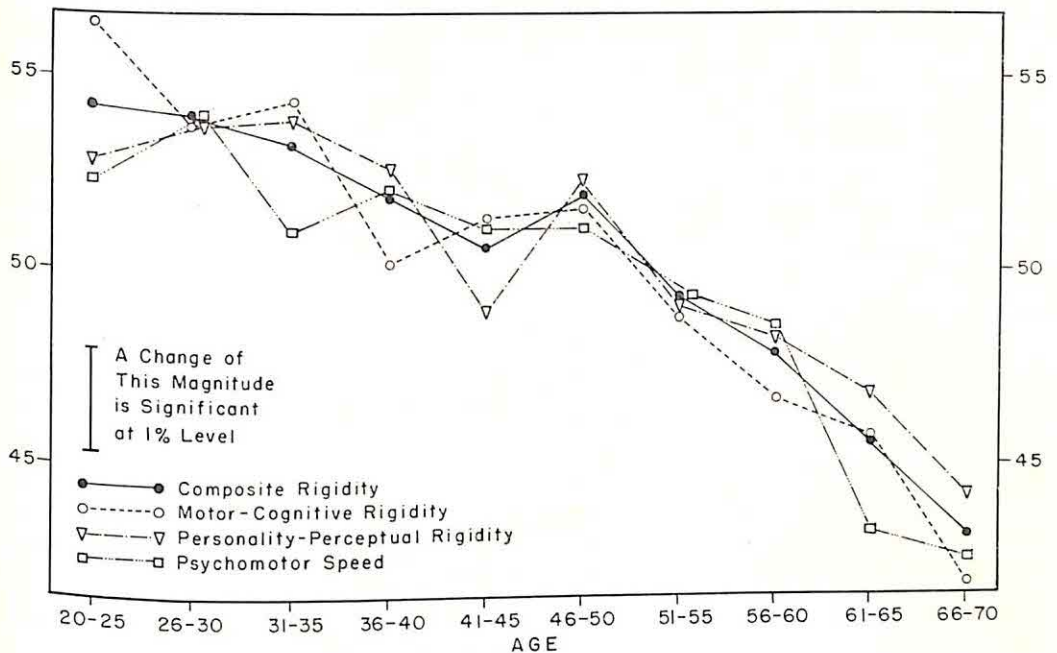


FIG. 1. Age changes in the measures of rigid behavior.

plotted for all groups. These are shown in Table 3 and Figure 2. It appears that the most dramatic decline is found on the motor-cognitive rigidity dimension, while the least (though still highly significant) change occurs on the personality-perceptual dimension. Significant decrements at the 1% level of confidence prevail for the motor-cognitive rigidity dimension already for the 26-30 year old group. On the personality-perceptual rigidity dimension, however, the decrement is not significant until the 41-45 year old group is reached. On the psycho-

TABLE 2
AGE CHANGES IN THE COMPOSITE RIGIDITY, MOTOR-COGNITIVE RIGIDITY,
PERSONALITY-PERCEPTUAL RIGIDITY, AND PSYCHOMOTOR SPEED SCORES

Age	Composite rigidity		Motor-cognitive rigidity		Personality-perceptual rigidity		Psychomotor speed	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
20-25	54.08	4.39	56.48	5.10	52.78	7.25	52.24	7.60
26-30	53.88	4.93	53.66	5.20	53.60	6.77	53.94	9.32
31-35	53.14	4.68	54.30	4.95	53.68	7.27	50.94	7.66
36-40	51.68	6.04	50.04	7.42	52.52	7.77	51.98	9.91
41-45	50.54	5.39	51.20	6.25	48.82	7.43	51.02	7.61
46-50	51.88	5.25	51.52	7.25	52.40	6.96	51.00	7.95
51-55	49.26	5.26	48.72	7.72	49.06	7.22	49.32	8.16
56-60	47.90	5.84	46.66	7.39	48.30	8.85	48.53	8.85
61-65	45.50	4.78	45.62	7.71	46.86	6.40	43.44	7.37
66-70	43.10	5.50	41.88	7.77	44.14	6.38	43.00	7.57

Note.—A high score on the rigidity variables means that this is a score in the *flexible* direction. It should also be noted that the statistics in this and all subsequent tables and graphs are based on a subgroup *N* of 50 for each group, consisting of 25 men and 25 women.

TABLE 3

ABSOLUTE MEAN DECLINE IN FLEXIBILITY WITH AGE IN *T*-SCORE POINTS FROM THE MEAN SCORE OF THE MOST FLEXIBLE SUBGROUP

Age	Com- posite rigidity	Motor- cognitive rigidity	Personal- ity-per- ceptual rigidity	Psycho- motor speed
20-25	—	—	0.90	1.70
26-30	0.20	2.82	0.08	—
31-35	0.94	2.18	—	3.00
36-40	2.40	6.44	1.16	1.96
41-45	3.54	5.28	4.86	2.92
46-50	2.20	4.96	1.28	2.94
51-55	4.82	7.76	4.62	4.62
56-60	6.18	9.82	5.38	5.40
61-65	8.58	10.86	6.82	10.50
66-70	10.98	14.60	9.54	10.94

motor speed dimension, significant decrement is noted at age 31-35; and finally, on the composite rigidity scores, significant decrement is noted beginning with the 36-40 year old group.

From these data it would appear that loss of flexibility would become noticeable in the forties and fifties, and that by the fifties an average loss of approximately one population *SD* has occurred. In terms of the dimensions tested it appears then that, whether by maturational change or environmental effect, there is a progressive and significant decrement with age in the ability of the average adult to react to environmental change.

Correlations were also computed, sep-

arately for each age group, among the factor scores and between the factor scores and the composite score. These and all subsequent correlations shown were computed by means of IBM equipment using a modification of Schaie's standard score method (18, 23). Table 4 does not seem to show any significant age trends in these intercorrelations.

2. *Age changes in mental abilities.* All PMA raw scores were converted into *T* scores to make them comparable to the rigidity scores, again using the mean and *SD* of the whole sample as the point of reference. A composite equivalent score was obtained by use of Thurstone's formula (already described) and converted into *T*-score form. Table 5 lists the means and *SD*s for each of the five mental abilities which were measured as well as for the composite IQ scores. Figure 3 shows the age gradients which result when the data in Table 5 are plotted. The ages given on these graphs are the midpoints of the subgroups.

Age changes on all five mental abilities, as well as on the composite IQ, are found to be significant at the 1% level of confidence. It will be noted, however, that the form of these gradients differs among the individual abilities. The peak level of performance is, with the exception of Reasoning, found at a later level

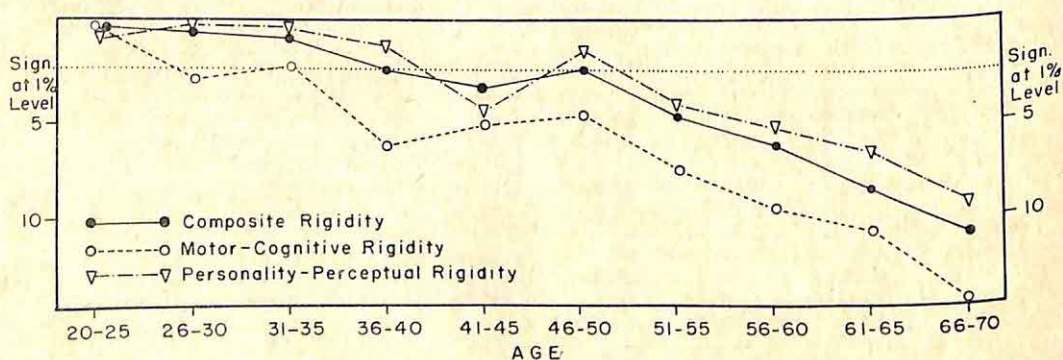


FIG. 2. Mean decrement in flexibility on all rigidity measures (in *T*-score points from mean of most flexible subgroup).

TABLE 4

INTERCORRELATIONS FOR THE FACTOR SCORES ON THE TEST OF BEHAVIORAL RIGIDITY AND THEIR CORRELATION WITH THE COMPOSITE RIGIDITY SCORE

Age	MCR-PPR	MCR-PS	PPR-PS	Comp-MCR	Comp-PPR	Comp-PS
20-25	— .04	.20	.24	.48	.68	.78
26-30	.24	.30	.13	.05	.61	.79
31-35	.06	.17	.35	.47	.74	.80
36-40	.18	.36	.26	.69	.65	.80
41-45	.21	.47	.36	.70	.73	.82
46-50	.04	.39	.39	.67	.64	.85
51-55	.10	.34	.10	.71	.56	.74
56-60	.32	.08	.31	.62	.79	.69
61-65	— .01	.50	.01	.77	.44	.79
66-70	.24	.38	.42	.74	.70	.80

than is suggested by comparable studies with other intelligence tests. On the composite IQ score, the peak level is reached by the 31-35 year old group. The peak levels of performance for the individual abilities are: Verbal-Meaning, 31-35 years; Space, 26-30 years; Reasoning, 20-25 years; Number, 46-50 years; and Word-Fluency, 31-35 years.

These findings indicate a wide range in the age trends for specific mental abilities. Since all the peak levels appear in the period from 20 to 50 years, this differential rise and decline of abilities might help to explain the tendency for global intelligence scores, when plotted against age, to show minimally significant decline for the period of the middle years and then, about age 50, to show

steeper decrements.

The absolute decrement in *T*-score points from the mean peak levels for each of these abilities is shown in Table 6. Figure 4 also presents this comparison of differential age changes graphically. It may be seen that there is variability not only in age of peak performance, but also in the age level at which a significant decrement from the peak level is first noticed. Thus it may be noted that, for the composite IQ score, a significant decrement is first observed for the 36-40 year old group. The ages for the individual mental abilities at which significant decrements are first seen are: Verbal-Meaning, Space, Reasoning and Word-Fluency, 36-40 years; while for Number it does not occur until age 56-60.

TABLE 5

AGE CHANGES ON THE PRIMARY MENTAL ABILITIES

Age	Verbal-Meaning		Space		Reasoning		Number		Word-Fluency		Composite IQ Score	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
20-25	51.40	8.15	54.80	11.38	56.24	8.82	47.52	9.89	50.05	9.37	51.90	9.46
26-30	55.35	8.10	55.93	9.93	55.93	8.53	50.57	9.15	51.10	10.44	54.25	9.18
31-35	55.50	8.13	55.27	9.38	55.39	8.99	52.37	9.83	53.69	8.56	55.35	9.23
36-40	52.82	11.20	52.97	7.93	52.45	9.74	51.33	10.71	50.62	8.93	52.43	9.15
41-45	52.57	8.49	50.99	9.38	53.02	7.95	52.34	9.31	52.78	9.74	53.08	8.07
46-50	52.29	9.17	50.07	9.82	50.30	9.17	53.26	10.38	52.37	9.60	52.39	9.54
51-55	49.13	8.79	47.79	9.40	48.39	9.10	52.12	10.00	51.25	9.51	50.10	8.96
56-60	46.61	9.76	45.23	10.58	45.39	8.36	50.11	8.77	49.25	9.37	47.04	8.70
61-65	43.03	8.21	45.40	8.51	42.05	7.10	45.08	8.90	45.23	11.14	42.55	8.31
66-70	41.34	7.93	41.60	7.48	40.88	5.86	45.30	8.75	43.70	8.21	40.88	7.44

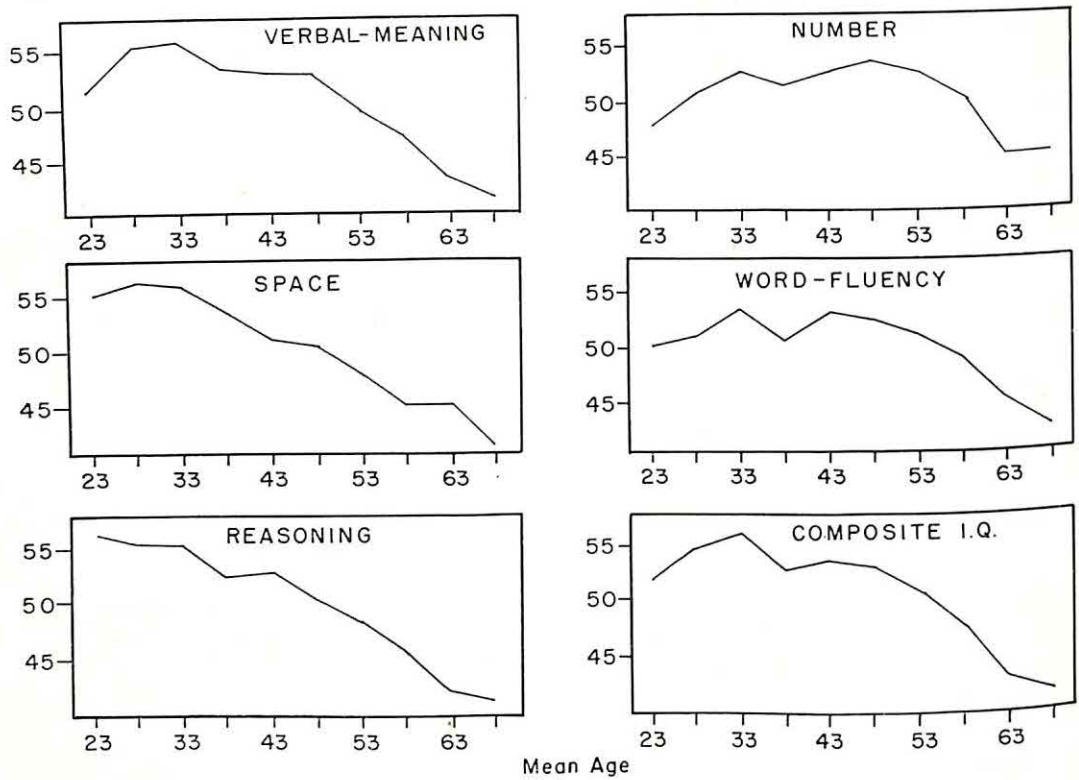


FIG. 3. Age changes in the Primary Mental Abilities.

More variation is shown in the rate of decrement. If decrements are assumed to become behaviorally noticeable if they have a magnitude of 5 points ($\frac{1}{2}$ SD) and marked if they exceed 10 points (1 SD), then these findings would indicate that such decline becomes *noticeable* at the

following age levels: Verbal-Meaning, 51-55; Space and Reasoning, 46-50; Number and Word-Fluency, 61-65; and total IQ, 51-55. Decrement in the mental abilities according to above criterion would be considered *marked* at the following ages: Total IQ, 61-65; Space and Reason-

TABLE 6
ABSOLUTE MEAN DECREMENT IN T-SCORE POINTS FOR THE PRIMARY MENTAL ABILITIES FROM MEAN PEAK LEVEL

Age	V	S	R	N	W	IQ
20-25	4.10*	1.13	—	5.74*	3.64*	3.45*
26-30	0.15	—	0.31	2.60*	2.59	1.10
31-35	—	0.66	0.85	0.80	—	—
36-40	2.68*	2.96*	3.79*	1.93	3.07*	2.92*
41-45	2.93*	4.94*	3.22*	0.92	0.91	2.27
46-50	3.21*	5.86*	5.94*	—	1.32	2.96*
51-55	6.37*	8.14*	7.85*	1.14	2.44	5.25*
56-60	8.86*	10.70*	10.85*	3.15*	4.44*	8.31*
61-65	12.47*	10.53*	14.19*	8.18*	8.46*	12.80*
66-70	14.16*	14.33*	15.36*	7.96*	9.99*	14.47*

* Significant at the 1% level of confidence.

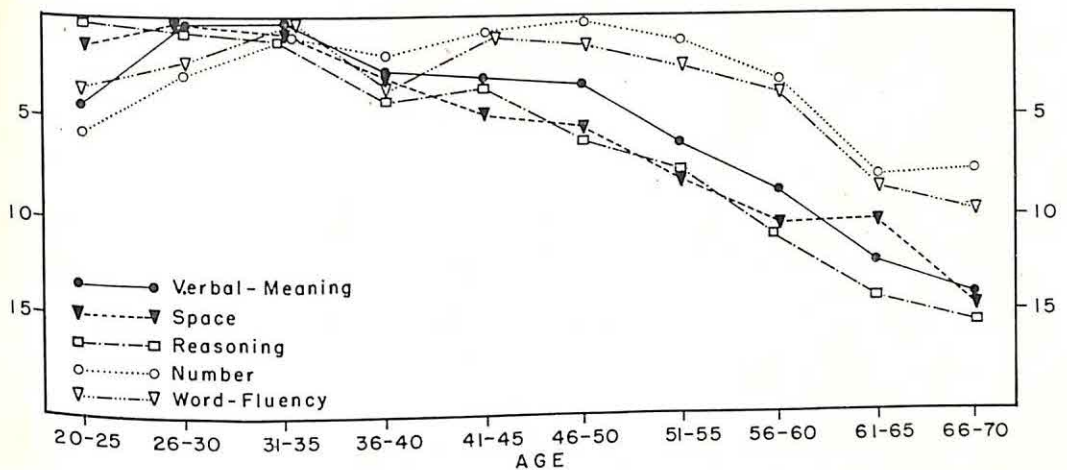


FIG. 4. Mean decrement in the Primary Mental Abilities from mean peak levels (in *T*-score points). (Significant at 1% level = 2.63; at 5% level = 1.98.)

ing, 56-60; Verbal-Meaning, 61-65; and Number and Word-Fluency probably in the seventies, if these age trends continue systematically.

Another matter of interest is the degree of variability shown by individuals among their different mental abilities. So that variability could be discerned, it was decided to compute for each *S* the variance of his PMA scores about his own mean. These variances were then normalized and transformed into *T* scores. Table 7 gives the mean variability for the different age groups. This table shows that there is a slight but significant tendency for intra-individual variability to increase with age and reach a peak at age 41-45. There is some decline in variability with the older groups, but this decline was not significant at the 1% level of confidence. These findings seem to agree to some extent with Balinsky's factorial study of the subtests of the Wechsler-Bellevue at various age levels, in which the variability was found to increase towards the middle years and then show decline (1). The peak level of variability was shown by Balinsky to be at age 25-29, however, and the decline in vari-

bility found by him was more marked than in the present study.

To shed more light on the problem of shifts in intra-individual differences with age, and the possible relationship of such changes to the "rigidity-flexibility" model, subtest intercorrelations were computed. Table 8 gives the intercorrelations of the subtests separately for each age group, and Table 9 lists the correlations between the individual mental abilities and the composite score.

If one considers average intercorrelations for each age level, the lowest relation is found at ages 36-45; but other shifts are not systematic enough to allow

TABLE 7
AGE CHANGES IN INTRA-INDIVIDUAL
VARIABILITY ON THE PMA

Age	Mean	SD
20-25	48.99	6.91
26-30	48.62	7.60
31-35	48.28	7.17
36-40	49.99	10.16
41-45	51.68	11.19
46-50	48.95	9.56
51-55	49.85	11.18
56-60	50.60	11.84
61-65	50.49	10.73
66-70	50.50	9.59

TABLE 8
INTERCORRELATIONS OF THE PRIMARY MENTAL ABILITIES

Age	V-S	V-R	V-N	V-W	S-R	S-N	S-W	R-N	R-W	N-W	Average
20-25	.28	.50	.47	.45	.42	.39	.43	.56	.58	.64	.472
26-30	.47	.71	.50	.46	.53	.39	.35	.59	.51	.42	.493
31-35	.49	.75	.57	.45	.46	.57	.18	.61	.52	.45	.505
36-40	.20	.51	.47	.53	.21	.09	.02	.56	.46	.38	.343
41-45	.21	.52	.37	.58	.39	.33	.18	.52	.33	.11	.354
46-50	.33	.71	.43	.64	.49	.36	.39	.44	.50	.44	.473
51-55	.45	.69	.31	.44	.56	.27	.39	.53	.39	.31	.434
56-60	.15	.66	.49	.53	.24	.18	.30	.42	.54	.53	.404
61-65	.31	.55	.41	.47	.33	.40	.38	.30	.54	.39	.408
66-70	.13	.54	.53	.26	.35	.49	.16	.80	.48	.32	.406

further conclusions. No systematic shifts are noted in the relationship between the individual abilities and in the composite IQ score.

The Relation Between the Changes in Behavioral Rigidity and Mental Abilities: Results of the Correlation Analysis

After considering the age changes in both rigidity and mental ability measures, the next step in the analysis of the experimental evidence is concerned with changes in the relationship between the two sets of measures. Correlations were therefore computed between the composite rigidity score and the PMAs as well as between the individual TBR factor scores and the PMAs.

The reader should be reminded here of the hypotheses submitted earlier. Hypothesis I stated that:

Behavioral age changes are a function of the individual's lack of ability to adjust to change and can thus be predicted for an advanced age from a knowledge of his standing on the rigidity-flexibility dimension at the basal age.

If this hypothesis were correct, the correlations between the TBR and the PMA would be expected to increase systematically with age. Hypothesis II stated that:

Behavioral age changes occur concomitantly with decreasing adjustment to change and the level of function for individuals at any given age level is therefore predictable from knowledge of [either] . . . dimension.

If this hypothesis were true one would expect to find significant but constant correlation throughout the adult age range, with possibly slightly decreasing values towards the upper end of the range. Hypothesis III stated that:

Behavioral age changes and changes in rigidity-flexibility show no significant relation whatsoever.

In this case insignificant correlation throughout the adult age range would be expected. Hypothesis IV, finally, is the obverse of Hypothesis I. It stated that:

Lack of ability of an individual to adjust to change may be a function of some other behavior variable and may be predictable for his old age from a knowledge of that variable at the basal age.

Again correlations between PMA and TBR would be expected to rise with age and the direction of the relation would

TABLE 9
CORRELATIONS BETWEEN THE PRIMARY MENTAL ABILITIES AND THE COMPOSITE IQ SCORE

Age	V	S	R	N	W
20-25	.66	.64	.79	.86	.82
26-30	.77	.67	.85	.81	.72
31-35	.80	.67	.84	.88	.67
36-40	.77	.32	.80	.82	.68
41-45	.74	.57	.78	.76	.61
46-50	.78	.63	.80	.79	.78
51-55	.72	.66	.84	.76	.67
56-60	.77	.47	.77	.79	.81
61-65	.71	.62	.71	.77	.70
66-70	.67	.56	.88	.90	.61

TABLE 10

CORRELATIONS BETWEEN THE COMPOSITE RIGIDITY SCORE AND THE PRIMARY MENTAL ABILITIES

Age	V	S	R	N	W	IQ
20-25	.50	.26	.63	.42	.55	.61
26-30	.67	.35	.68	.56	.50	.71
31-35	.69	.39	.52	.51	.40	.65
36-40	.60	.27	.63	.67	.52	.80
41-45	.75	.27	.52	.44	.48	.69
46-50	.71	.25	.53	.48	.60	.67
51-55	.58	.37	.58	.29	.54	.61
56-60	.49	.15	.56	.36	.63	.60
61-65	.35	.00	.44	.11	.32	.32
66-70	.59	.18	.47	.44	.38	.66

TABLE 12

CORRELATIONS BETWEEN THE PERSONALITY-PERCEPTUAL RIGIDITY SCORE AND THE MENTAL ABILITIES

Age	V	S	R	N	W	IQ
20-25	.31	.06	.20	.08	.24	.21
26-30	.31	.02	.24	.10	.12	.19
31-35	.42	.21	.27	.25	.25	.34
36-40	.30	.20	.30	.37	.27	.42
41-45	.46	.23	.39	.18	.23	.40
46-50	.36	-.09	.17	.31	.16	.27
51-55	.12	.19	.18	.05	.14	.16
56-60	.21	.13	.32	.16	.36	.32
61-65	-.12	-.14	-.08	-.06	-.09	-.12
66-70	.18	.05	.09	.08	-.02	.10

be testable by means of the analysis of covariance.

Tables 10-13 give the intercorrelations between the composite rigidity score and the TBR factor scores with the composite IQ as well as the individual PMA scores, and Figure 5 shows a graphic representation of the age gradients for the correlations between the rigidity scores and the composite IQ measure.

Figure 5, as well as the data in Tables 10-13, indicate that there is a general tendency for these correlations to be significant and to be relatively constant over the adult age range, except for a slight drop in the older age groups. Correlations for samples of the magnitude here used are significant at the 5% level of confidence if they exceed .27 and at

the 1% level of confidence if they exceed .35.

While the changes in the relation between rigidity and the PMAs are fairly consistent, they are by no means uniform for the different rigidity dimensions or the individual mental abilities. It should be noted that the personality-perceptual rigidity score shows lower correlation with the PMAs throughout, and at the top end of the age range seems to have no relation at all to the global intelligence score. This is in contrast to the motor-cognitive rigidity and the psychomotor speed scores, which correlate significantly with the PMAs throughout.

Since the composite rigidity score appears to show, on the whole, the highest correlations with the individual PMA

TABLE 11

CORRELATIONS BETWEEN THE MOTOR-COGNITIVE RIGIDITY SCORE AND THE PRIMARY MENTAL ABILITIES

Age	V	S	R	N	W	IQ
20-25	.18	.31	.31	.27	.32	.37
26-30	.53	.42	.63	.33	.25	.53
31-35	.40	.56	.30	.40	.13	.44
36-40	.39	.20	.47	.47	.29	.54
41-45	.54	.25	.28	.38	.33	.52
46-50	.60	.36	.46	.24	.61	.56
51-55	.54	.42	.53	.22	.31	.51
56-60	.18	.04	.30	.05	.27	.22
61-65	.34	.14	.47	.11	.41	.38
66-70	.60	.19	.35	.35	.25	.56

TABLE 13

CORRELATIONS BETWEEN THE PSYCHOMOTOR SPEED SCORE AND THE PRIMARY MENTAL ABILITIES

Age	V	S	R	N	W	IQ
20-25	.46	.18	.70	.49	.50	.61
26-30	.56	.32	.55	.62	.57	.70
31-35	.58	.18	.50	.41	.54	.56
36-40	.57	.19	.57	.58	.54	.72
41-45	.68	.14	.49	.44	.51	.65
46-50	.56	.25	.50	.47	.51	.61
51-55	.49	.14	.47	.32	.61	.55
56-60	.62	.14	.56	.48	.68	.68
61-65	.45	.02	.48	.14	.30	.36
66-70	.52	.14	.58	.52	.52	.64

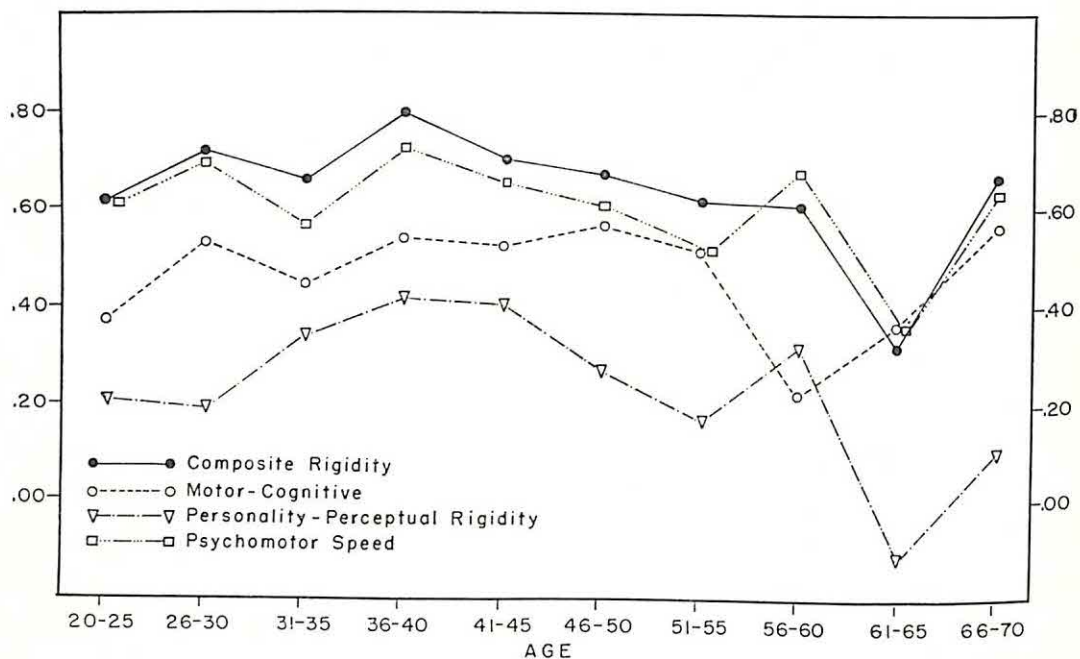


FIG. 5. Correlation between TBR scores and PMA composite IQ.

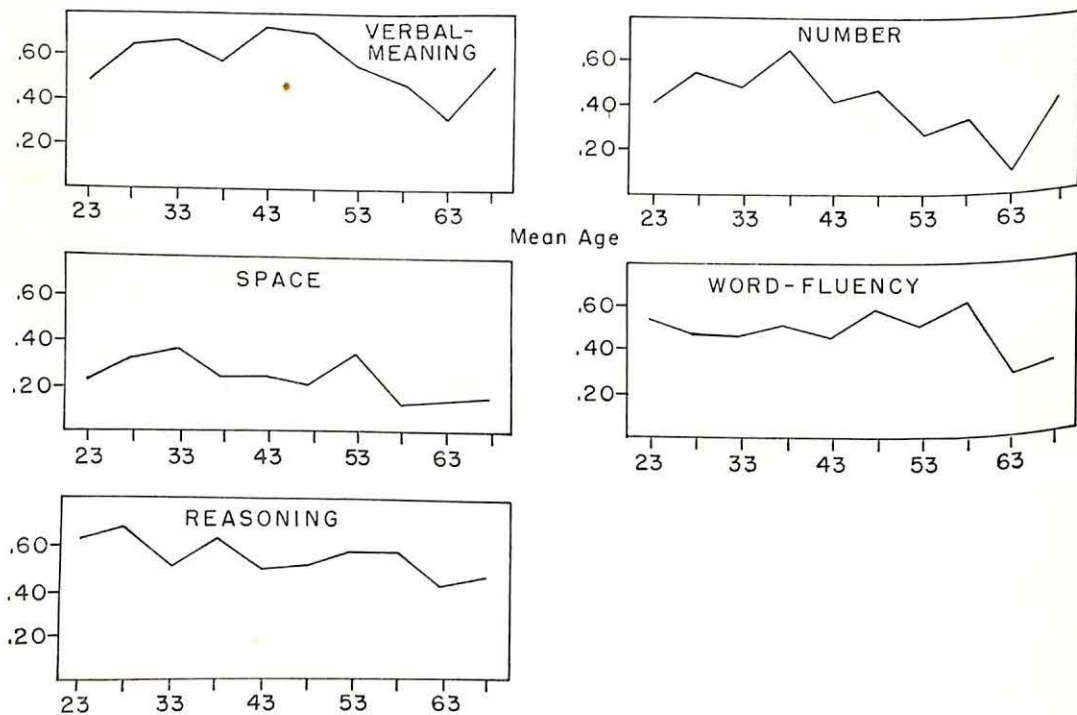


FIG. 6. Correlation between the PMAs and the Composite Rigidity score.

TABLE 14
CORRELATION OF YEARS OF EDUCATION WITH THE TBR AND PMA SCORES

Age	MCR	PPR	PS	COMP	V	S	R	N	W	IQ
20-25	.03	.24	.33	.35	.45	.22	.20	.35	.23	.40
26-30	.37	.16	.47	.50	.51	.27	.59	.38	.25	.51
31-35	.14	.27	.50	.48	.50	.16	.34	.40	.22	.42
36-40	.51	.34	.52	.64	.66	.16	.47	.50	.59	.69
41-45	.44	.13	.30	.38	.24	.11	.20	.04	.16	.19
46-50	.36	.23	.47	.51	.59	.19	.40	.13	.40	.41
51-55	.44	.04	.72	.62	.63	.16	.53	.37	.45	.58
56-60	.17	.38	.60	.56	.43	-.01	.58	.43	.48	.54
61-65	.30	.04	.54	.44	.40	.10	.40	.21	.26	.36
66-70	.38	.19	.52	.54	.54	.19	.56	.48	.37	.59

scores, its relationship with each of the individual mental abilities has been plotted in Figure 6. It will be noted that the correlations are again significant over the entire age range, except for Space, with some apparent decrease in the correlations with age. The slopes and levels of these age gradients, however, differ among the abilities, with little or no significant drop in the correlations with Reasoning and Word-Fluency, fairly marked drop for Verbal-Meaning and Number, and relatively low or insignificant correlation with Space.

On the basis of the preceding results it appears that Hypotheses I and IV are not tenable, while, with the possible exception of the relationship with Space, evidence favorable to Hypothesis II has been presented. A more formal statistical test of these hypotheses will be presented in the next section concerned with the results of the analyses of variance and covariance.

Before proceeding further, some of these relationships should be re-examined with respect to another variable. It was noted in the description of the sample that there were some tendencies for the older age groups to show a somewhat lower mean level of education as is characteristic of the general population. Since this factor is generally

credited with confounding the results of aging studies, it was felt appropriate to analyze the influence of amount of education. Correlations were therefore computed between years of formal education and the PMA and TBR scores. These are listed in Table 14.

It is apparent from Table 14 that years of education are significantly related to most of the variables studied, but few or no significant age trends are noted. Education seems to be related most significantly with the psychomotor speed score. One would infer from this observation that one of the principal effects of added years of education might be to add to one's ability to emit familiar responses without hesitation. Education seems thus more important in facilitating the handling of known data than in influencing a person's ability to handle new learning or unfamiliar situations. There is, however, a mild correlation between the measures of motor-cognitive and personality-perceptual rigidity and education. Education shows, as would be expected, moderately positive correlation with the composite IQ scores.

The next step in the analysis of the influence of education was to compute partial correlation coefficients between the TBR scores and the composite IQ with the effect of education held constant.

TABLE 15

CORRELATIONS BETWEEN THE COMPOSITE IQ AND THE TBR SCORES ADJUSTED FOR THE EFFECT OF YEARS OF EDUCATION

Age	MCR	PPR	PS	COMP
20-25	.39	.13	.56	.55
26-30	.43	.13	.61	.61
31-35	.45	.26	.44	.57
36-40	.30	.27	.58	.64
41-45	.49	.38	.63	.68
46-50	.49	.19	.51	.55
51-55	.35	.17	.24	.40
56-60	.16	.15	.54	.56
61-65	.31	-.15	.21	.20
66-70	.45	-.02	.47	.50

Table 15 lists these partial correlations and Figure 7 shows the new age gradients plotted with these adjusted correlation values. It can be seen that adjusting for the effect of differences in years of education merely tends to reduce all correlations somewhat but has little or no systematic effect on the slopes of the age gradients.

The Test of the Principal Hypotheses: Results of the Analysis of Variance and Covariance.

The final part of this section is con-

cerned with the formal statistical test of the hypotheses by means of the analysis of variance and covariance. The methods used are those described by Edwards (5).

The analysis of variance was used as a first step to test the null hypothesis with respect to the significance of the differences between age groups on both the TBR and PMA scores. Since the *T* scores were used for these computations, the sums of squares for the different variables, as summarized in Table 16, are directly comparable. All *F* ratios are, of course, found to be highly significant.

The next step in this analysis is concerned with the test of the hypotheses suggested by the rigidity-flexibility model. If Case I is true, one should find that age changes on the PMA are reduced to nonsignificance when the means are adjusted for level of rigidity, since the hypothesis tested proposes that age changes in intellectual abilities are a function of the differences in level of rigidity. The null hypothesis states that if changes in intelligence can be accounted for on the basis of changes in rigidity, then the *F* ratio between the ad-

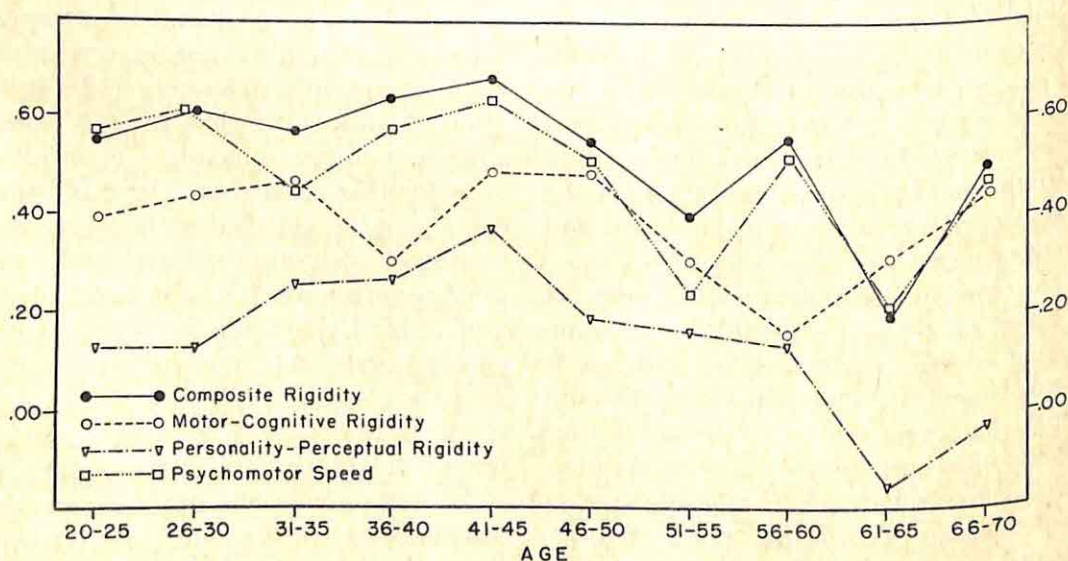


FIG. 7. Correlation between TBR scores and PMA composite IQ (effect of education partialled out).

TABLE 16

RESULTS OF THE ANALYSIS OF VARIANCE TESTING
THE SIGNIFICANCE OF THE DIFFERENCES
BETWEEN AGE GROUPS ON THE TBR
AND PMA SCORES

Variable	Between-group variance	Within-group variance	F ratio ^a
Composite rigidity	671.86	27.95	24.04
Motor-cognitive rigidity	575.19	46.83	12.28
Personality-perceptual rigidity	529.11	53.88	9.82
Psychomotor speed	666.20	68.76	9.69
Composite IQ	1,215.81	79.57	15.28
Verbal-Meaning	1,202.55	79.91	15.05
Space	1,220.22	91.11	13.60
Reasoning	1,603.27	72.61	22.08
Number	448.70	93.95	4.78
Word-Fluency	517.37	92.61	5.59

^a With 9 and 490 degrees of freedom respectively, an *F* ratio will be significant at the 1% level of confidence if it exceeds 2.45.

justed variances should be a chance deviation from zero. Table 17 summarizes the results of the analysis of covariance for the regression of the composite IQ upon the TBR composite and factor scores and also, as a reverse test of possible evidence for the validity of Case IV, for the regression of the composite IQ score on the TBR composite score. With 9 and 489 degrees of freedom respectively, an *F* ratio is significant at the 1% level of confidence if it exceeds 2.43. Every *F*

TABLE 17

RESULTS OF THE ANALYSIS OF COVARIANCE TESTING THE EFFECT OF THE REGRESSION OF THE COMPOSITE IQ SCORE UPON THE TBR SCORES AND OF THE COMPOSITE RIGIDITY UPON THE IQ SCORE

Variable	Adjusted between-group variance	Adjusted within-group variance	F ratio
Composite rigidity	150.24	48.61	3.09
Motor-cognitive rigidity	535.78	63.31	8.46
Personality-perc. rigidity	744.55	75.14	9.91
Psychomotor speed	847.88	49.52	17.12
Composite IQ	175.95	17.08	10.30

TABLE 18

RESULTS OF THE ANALYSIS OF COVARIANCE TESTING THE EFFECT OF THE REGRESSION OF THE INDIVIDUAL MENTAL ABILITIES UPON THE COMPOSITE RIGIDITY SCORE

Variable	Adjusted between-group variance	Adjusted within-group variance	F ratio
Verbal-Meaning	190.10	51.52	3.69
Space	436.54	85.47	5.11
Reasoning	284.41	50.81	5.60
Number	315.51	76.54	4.12
Word-Fluency	166.30	71.29	2.39

ratio in Table 17 is therefore significant, and the null hypothesis as stated is not verified.

Comparison of Tables 16 and 17 shows that the maximum reduction in the *F* ratios, by adjusting the means for the effects of correlation, occurred for the composite rigidity score. As a further check, the analysis of covariance was therefore used to test the effect of the regression of the individual mental abilities upon the composite rigidity score. Table 18 lists the results of this analysis, which show reduction in the *F* ratios which, however, remain significant. The null hypothesis may thus also be rejected for the individual mental abilities.

As a result of the correlation analysis it appears that Case II is the more plausible one. Hypothesis II required that there be substantial correlation between rigidity-flexibility and intelligence at each age level and that prediction of intellectual functioning would be possible from a knowledge of rigidity scores. In order to obtain further evidence for the validity of the second alternative, the records for the *Ss* in each age group were rank ordered in terms of the composite rigidity score. The 10 most flexible and the 10 most rigid subjects in each age group were then selected and mean values were computed for these extremes on

TABLE 19
MEANS ON THE PRIMARY MENTAL ABILITIES FOR THE EXTREME FLEXIBLE AND
THE EXTREME RIGID MEMBERS OF THE SAMPLE

Age		20-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70
V	r	44.5	48.0	46.3	41.8	43.2	43.0	41.2	43.4	34.9	34.5
	f	57.1	61.3	60.5	58.8	61.5	60.5	55.8	54.8	44.3	48.5
S	r	53.8	51.5	48.8	51.0	47.9	48.4	43.0	44.2	44.2	38.3
	f	61.0	57.3	57.5	58.2	56.8	54.6	53.0	49.0	42.7	40.0
R	r	49.4	49.8	47.9	44.2	49.0	42.4	42.0	39.2	37.3	36.9
	f	65.6	63.8	59.0	61.5	58.6	54.8	56.1	52.0	47.5	43.5
N	r	42.8	40.8	42.0	41.0	47.4	46.1	48.8	48.9	43.3	39.1
	f	57.9	57.5	54.0	61.0	56.4	60.9	58.4	56.8	47.0	50.9
W	r	46.0	43.2	46.4	41.7	46.6	43.4	45.1	41.0	37.9	41.7
	f	61.2	53.4	56.5	52.8	58.9	58.8	58.5	58.4	52.1	46.4
IQ	r	45.7	46.4	44.6	41.5	46.2	43.2	43.1	42.0	36.9	35.0
	f	63.4	60.0	59.0	61.2	59.6	61.0	58.7	56.0	46.2	45.6

the measures of mental ability.

While there are some age trends in the differences between the flexible and rigid groups, these differences appear to be significant at all ages and for all abilities. Differences range from one-half to two population *SDs*. When mean differences are computed for the extremes for the total population, they are found to be in excess of 1.0 *SD*, except for Space which shows a mean difference of 0.6 *SD*. Table 19 lists the mean values for the extreme flexible (f) and extreme rigid (r) subgroups, classified by age.

Table 20 lists the absolute differences in *T*-score points, all of which are significant at the 1% level of confidence,

except the differences for the two oldest age groups on Space. There is a tendency for all these differences to be more pronounced in the middle years, with increasing convergence towards the upper end of the age continuum. It appears that, due to the fact of general decrement in abilities, the range of talent becomes more restricted in the older age groups. However, significant differences in the various abilities are maintained between the rigid and flexible individuals throughout the age range. This evidence would seem to favor Hypothesis II and lead to a rejection of the null hypothesis with respect to Hypotheses I, III, and IV.

TABLE 20
DIFFERENCES IN THE MEAN SCORES ON THE PRIMARY MENTAL ABILITIES FOR THE EXTREME
RIGID AND FLEXIBLE GROUPS (IN *T*-SCORE POINTS)

Age	20-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70
V	12.6	13.3	14.3	17.0	18.3	17.5	14.5	11.4	9.4	14.0
S	7.2	5.9	8.7	7.2	8.9	6.2	10.0	4.8	1.5	1.7
R	16.2	14.0	11.1	17.3	9.6	12.4	14.1	12.8	10.2	6.6
N	15.1	16.7	12.0	20.0	9.0	14.8	9.6	8.0	3.6	11.8
W	15.3	10.3	10.1	11.1	12.3	15.4	13.4	17.4	14.2	4.8
IQ	17.7	13.6	14.5	19.7	13.3	17.8	15.6	14.0	9.2	10.6

DISCUSSION

More than anything else, it would seem that the evidence on concomitant age changes both in the rigidity and intellectual measures gives support to the school of thought which sees decline as a maturational phenomenon which occurs at a relatively constant rate and whose degree is presumed to be a function of original endowment. In the case of the mental abilities, this interpretation would seem to be reinforced by the findings of decreasing response variability with increasing age.

The results of the present study suggest that both rigidity-flexibility and measures of mental ability may be useful to predict level of functioning at any given age level. There was little support, however, for the contention that either dimension may be used for the long-range prediction of specific age gradients in other behavior characteristics. While one cannot use rigidity-flexibility to predict absolute decline, the relationships found may be used to make predictions about the relative rank an individual is likely to attain at a later age.

Although measures of three distinct rigidity dimensions were used, it appears that a linear combination of these factors had the best predictive power and that such a composite score discriminated best on other variables. While there is a great deal of overlap about the mean of this score, it was found that at the extremes flexible individuals excelled the rigid with respect to many behavior and personal characteristics.⁷ This finding seems to imply that a composite rigidity measure, combining estimates of the several

dimensions used to define rigidity, is a useful and discriminating tool.

The rigidity-flexibility dimension was found to be of almost equally discriminative power throughout the adult age range, even though restriction of the range of observed rigidity measures occurred at the upper age levels. This finding might be interpreted by suggesting that the rigidity measures may give an indication of how well an individual is able to make use of his capacities. Thus an individual who scores high on the TBR probably makes effective use of his capacities, while one who scores low does not. No direct evidence on this point is given in this study, however, and one would probably have to define independent indices of optimal functioning to be correlated with our rigidity measures for further evidence.

It was noted that, among the rigidity factor scores, the ones showing highest relations with other measures were the motor-cognitive rigidity and the psychomotor speed scores. These scores presumably are indices of the difficulty in adjusting to change in familiar tasks and the ability to emit familiar responses and were expected to correlate best with the measures of cognitive behavior. The personality-perceptual rigidity score showed only low correlation with the cognitive material, but its inclusion in the composite score raised the predictive power of the latter.

Because of the high correlations between measures of intelligence and of rigidity an analysis of the predictive power of extreme scores was attempted also in terms of the composite intelligence measures. This analysis showed that discrimination occurred also at the younger ages, but convergence of the mean flexibility-rigidity scores for the

⁷ The extreme flexible subjects were found to exceed the extreme rigid in a positive direction in education, income, occupational status, social responsibility, and other variables (24).

extreme low and high intelligence groups was seen in middle age. This finding would seem to be further evidence against Hypothesis IV and would suggest that at least in the older age groups the rigidity-flexibility scores are better indicators of effective functioning and have more predictive power than performance on the intelligence test. The results of these extreme splits for both rigidity-flexibility and intelligence suggest that optimal functioning at young maturity and during the middle years may involve the interaction of both flexibility and intellectual level, while at the older ages maintenance of flexibility increases in importance.

In terms of the long-range prediction of behavior, it seems safe to conclude that this study presents additional evidence to show significant age decrements in important aspects of behavior. Considerable differences are also shown to exist among different measures as to their peak age of performance, the slope of the decrement gradient, and the variability of individuals at given age levels. It is of considerable interest to note that the peak levels found in the present study come at later ages than those found in the cross-sectional studies of the 1930's. Not enough time seems to have elapsed, however, to attribute these differences entirely to cultural changes, and the different tests used may be at least partially responsible for these findings.

On the basis of these findings with respect to differing peak ages for different abilities, it may be suggested that the adult plateau shown by global measures of intelligence would seem to be caused by the fact that during this period growth still occurs in some abilities while decline has already set in with others. The point where steep and behaviorally significant decrement begins occurs at

that age where growth has ceased in all the abilities which are measured to yield the intelligence estimate. Likewise it may be suggested that the findings of different variability at given ages, for different measures of ability, may account for the apparent differences in the rate of decline of individuals. These differences may well be a function of their attaining a peak level of development at different points.

From the age gradients presented for both the measures of rigidity and mental ability it would appear that, in general, an individual's status at a given age will be a function of his original peak level. Knowing the individual's basal level and the average age changes observed in the population, one might then be able to make a rough prediction as to what his level of performance is likely to be at a later age. Further work is needed, however, to determine the effect of differential basal levels.

SUMMARY

1. An analysis was made of studies of developmental changes in adult behavior, and explanatory models using the concept of behavioral rigidity were considered.

2. Alternative models for the relationship between developmental changes in intelligence (as measured by the Primary Mental Abilities) and measures of rigid behavior were formulated. Case I stated that maintenance of intellectual ability is a function of "flexibility," in which case a positive and systematically increasing relationship between measures of intelligence and rigidity-flexibility would occur with increasing age. Case II stated that both intellectual ability and behavioral rigidity show concomitant age changes, in which case either might be predicted by the other, and that a high

positive relationship should be found throughout the adult age span, with somewhat decreasing correlations due to the restriction of range in the older age groups. Case III stated that, while age changes in the mental abilities and rigidity-flexibility occur, they would show no systematic relation and zero or spurious correlation would be expected. Case IV finally involved the falsification of Case I, namely that maintenance of flexibility is a function of intellectual ability.

3. A sample of 500 adults ranging in age from 20 to 70 years, subdivided into five-year age groups, each containing 25 men and 25 women, was drawn from a pool of about 18,000 potential subjects who were members of a prepaid medical plan and were reasonably representative of a large portion of the socioeconomic continuum.

4. The subjects were tested with Thurstone's SRA Primary Mental Abilities Test and with Schaie's Test of Behavioral Rigidity. Information was also obtained on measures of family status and socioeconomic level.

5. Significant age differences were found on all measures of mental abilities and behavioral rigidity. There was considerable difference in the decrement shown on the different measures, in the ages of peak performance for the different abilities, in the changes in variability, and in the rate of decline.

6. Correlations between the measures of rigidity and mental ability and an analysis of variance and covariance were used to test the principal hypotheses. The correlations between the measures of rigidity and the mental abilities were found to be significantly high throughout the adult age span with a tendency to decrease towards the upper end of the age continuum. When the analysis of covariance was used to adjust the between-group differences on the mental abilities for the effect of rigidity, significant differences remained.

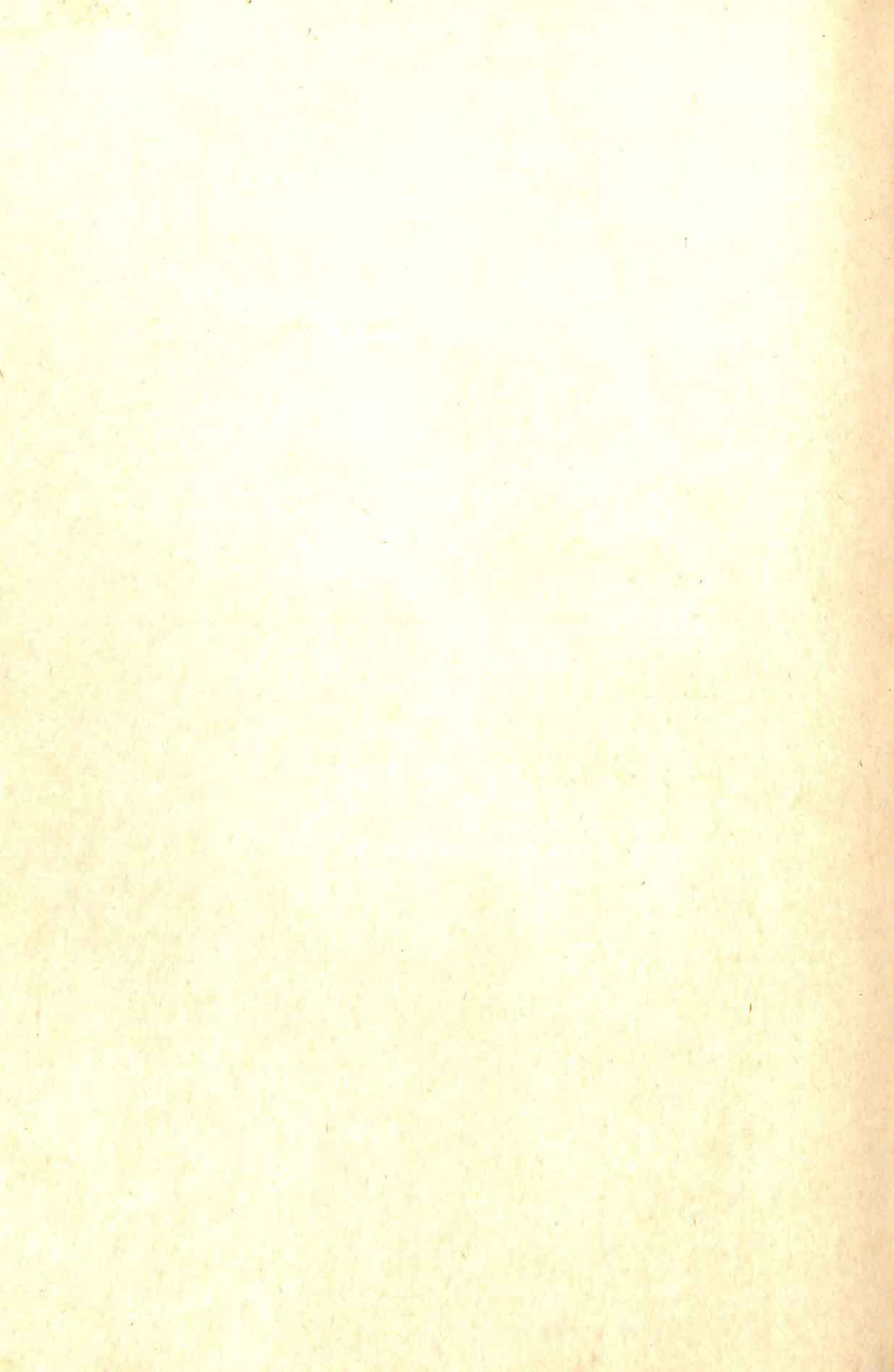
7. On the basis of this evidence neither Case I nor its obverse Case IV were verified, but Case II which favors interaction between rigidity-flexibility and intelligence at every age level seems to be supported.

REFERENCES

1. BALINSKY, B. An analysis of the mental factors of various age groups from nine to sixty. *Genet. Psychol. Monogr.*, 1941, 23, 191-234.
2. BERNSTEIN, E. Quickness and intelligence. *Brit. J. Psychol., Monogr. Suppl.*, 1924, 3, No. 7.
3. DAVIDSON, HELEN & KRUGLOW, LORRAINE. Characteristics of the institutionalized aged. *J. consult. Psychol.*, 1952, 16, 6-12.
4. DOERKEN, H. JR. Psychometric differences between senile dementia and normal senescent decline. *Canad. J. Psychol.*, 1954, 8, 187-196.
5. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
6. ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.
7. EYSENCK, M. The psychological aspects of aging and senility. *J. ment. Sci.*, 1946, 92, 172-181.
8. GOUGH, H. G. The C.P.I. Rigidity scale. Unpublished mimeographed manuscript, Institute of Personality Assessment and Research, Univer. of California, 1951.
9. GOUGH, H. G. *The California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1955.
10. GOUGH, H. G., McCLOSKEY, H., & MEEHL, P. E. A personality scale for social responsibility. *J. abnorm. soc. Psychol.*, 1952, 47, 73-80.
11. GUTTMAN, L. A technique for scale analysis. In *Measurement and prediction, the American soldier: Studies in social psychology in World War II. Vol. IV*. Princeton: Princeton Univer. Press, 1950.
12. HORST, P., & SCHAIE, K. W. The multiple group method of factor analysis and rota-

- tion to a simple structure hypothesis. *J. exper. Educ.*, 1956, 24, 231-237.
13. JACOBS, R. The reliability and intercorrelations of the scores on the SRA Primary Mental Abilities test. *Educ. Rec. Bull.*, 1948, 49, 49-59.
 14. LANKES, W. Perseveration. *Brit. J. Psychol.*, 1915, 7, 387-419.
 15. LEVINE, A. S. Perseveration, rigidity and persistence. *Psychol. Rep.*, 1955, 1, Monogr. Suppl. No. 5, 1-20.
 16. LEVINE, A. S. "Perseveration" or "the central factor." *Psychol. Rep.*, 1955, 1, Monogr. Suppl. No. 5, 247-265.
 17. LINDEN, M. E., & COURTNEY, D. The human life cycle and its interruptions: A psychological hypothesis: Studies in gerontologic human relations. *Amer. J. Psychiat.*, 1953, 109, 906-915.
 18. LUNNEBORG, C. E., WRIGHT, C. E., & AX, A. F. A set of statistical boards for the IBM type 604 electronic calculator. Unpublished mimeographed manuscript, Dept. of Psychiatry, Univer. of Washington, 1955.
 19. SCHAIE, K. W. Measuring behavioral rigidity: A factorial study of some tests of rigid behavior. Unpublished masters thesis, Univer. of Washington, 1953.
 20. SCHAIE, K. W. *Manual of instructions for the Test of Behavioral Rigidity, Form R-2*. Washington, D.C.: American Documentation Institute, Document No. 4651, 1955.
 21. SCHAIE, K. W. A test of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1955, 51, 604-610.
 22. SCHAIE, K. W. Some developmental concomitants of rigid behavior. Unpublished doctoral dissertation, Univer. of Washington, 1956.
 23. SCHAIE, K. W. The standard score method of computing intercorrelation matrices on the IBM type 604 electronic calculating punch. *IBM Technical Newsletter*, 1956, 12, 17-31.
 24. SCHAIE, K. W. Differences in some personal characteristics of "rigid" and "flexible" individuals. *J. Clin. Psychol.*, 1958, 14, 11-14.
 25. SCHAIE, K. W., & ROSENTHAL, F. Scale of Rigidity, Form LM-1. Unpublished adaptation of the CPI rigidity scale, Univer. of California, 1951.
 26. SCHAIE, K. W., ROSENTHAL, F., & PERLMAN, R. M. Differential deterioration of factorially "pure" mental abilities. *J. Geront.*, 1953, 8, 191-196.
 27. SCHEIER, I., & FERGUSON, G. A. Further factorial studies of tests of rigidity. *Canad. J. Psychol.*, 1952, 6, 18-30.
 28. SHOCK, N. W. Biology of aging. In T. L. Smith (Ed.), *Problems of America's aging population*. Gainesville: Univer. Florida Press, 1951.
 29. THURSTONE, L. L., & THURSTONE, T. G. *Factorial studies of intelligence*. Chicago: Univer. Chicago Press, 1941.
 30. THURSTONE, L. L., & THURSTONE, T. G. *Primary Mental Abilities Test*. Chicago: Science Research Associates, 1947.
 31. THURSTONE, L. L., & THURSTONE, T. G. *Examiner Manual for the SRA Primary Mental Abilities*. Chicago: Science Research Associates, 1949.
 32. WALKER, K. F., STAINES, R. C., & KENNA, J. C. p tests and the concept of mental inertia. *Charact. & Pers.*, 1943, 12, 32-43.
 33. WERNER, H. The concept of rigidity. *Psychol. Rev.*, 1946, 53, 42-53.

(Accepted for publication May 30, 1957)



Psychological Monographs: General and Applied

Attitude Change: Instability of Response and Acquisition of Experience¹

GEORGE KATONA

Survey Research Center, University of Michigan

SUPPOSE some people express different attitudes at two successive points of time; to what circumstances can the change be attributed? This is the major question raised in this paper. The same question may be asked regarding any two successive verbal responses, given by the same persons, whether they represent answers to questions of fact and information or expressed opinions and attitudes. Though some references will be made to other responses, the studies presented here are concerned mainly with short-run expectations regarding economic matters. Expectations may be defined as subjective notions of things to come and are therefore attitudes with a future time perspective (see 7). Most economic expectations differ not only from answers to questions of fact but also from deep-seated attitudes which tend to endure in spite of changing circumstances.

The starting point of the study is a differentiation between instability of response and acquisition of experience. The latter concept is then subdivided into personal experiences affecting one person individually, and acquisition of information which is available to very many people. We may summarize the findings obtained by distinguishing three major circumstances which may account for changes in successive responses.

Misclassifications. In a narrow sense this term includes clerical errors, interviewing errors, misunderstanding by respondents, and reporting errors. In a broader sense one may include in the same category effects of the mood of the respondent, and changes resulting from guessing or expressing ad hoc opinions when a respondent does not know the answer but is induced to reply.

Personal Experiences. In the typical learning situation, a person changes his response because he has learned something in the time period between the first and the second test. Events and developments affecting the respondent between the first and second interview (for instance, an accident or a promotion) are examples of personal experiences which may influence his opinions and attitudes.

These are instances of "true change" which are important for the individual. Yet such changes may be unimportant in the aggregate because they may cancel

¹The author is greatly indebted to the Ford Foundation as well as to his colleagues at the Survey Research Center who made the surveys reported here possible and helped him with analytical and statistical problems. Although it is difficult to single out a few people when a study is based on the collaboration of a rather extensive group of research workers, the author feels particularly indebted to William Hays, James N. Morgan, and Eva Mueller. Parts of this study, together with other findings derived from the Ford Foundation panel, were presented to a conference arranged by the National Bureau of Economic Research on The Quality and Economic Significance of Anticipations Data in Princeton, November 1957.

out. This assumption was confirmed, and the following generalization emerged: When individuals change their opinions and attitudes because of personal experiences, the distribution of opinions and attitudes in the entire sample is likely to vary little, because shifts from Attitude A to B and from Attitude B to A cancel out. More important is evidence supporting the validity of the same principle expressed in reverse: When two successive measurements of subjective notions yield similar aggregative distributions, then it is likely that many individuals have changed their attitudes in one direction, and many others in the opposite direction.

Social Learning. Acquisition of information by broad groups of people who receive the same information about new developments in their environment (for instance, receive widely circulated news about general price or wage increases) must be distinguished from personal experiences. When significant new developments take place in the environment and information about them is widely transmitted, many people comprehend the information in a similar manner, the information is reinforced through personal contacts and discussions, and similar changes in attitudes are induced. At the same time, "swimming against the current" becomes difficult even for people with contrary personal experiences. Therefore, when in the aggregate there is a substantial change in subjective notions, then it is likely that individual changes will be predominantly in one direction; the population tends to fall into two major groups: those who shifted in one direction and those who did not change at all. In other words, in this situation the total number of changers will be close to the minimum number

required to bring about the aggregative change in the distribution.

This report on the studies is organized as follows: The first two short sections are concerned with problems of definition and with statistical problems of measurement. The quantitative findings are presented in the third major section. This section presents descriptive results on the relation between aggregative changes in attitudes and changes by individuals, the findings regarding the origin of changes in expectations, and effects of attitude change on nonverbal behavior. In the last major section the implications of the findings are discussed. Implications for survey research and business cycle research are presented in some detail, while the closing words on implications for behavioral theory require elaboration aided by additional studies.

Differences and similarities in the changes of attitudes of individuals and of large groups of people—as, for instance, a representative sample of American families—are crucial for this study. Therefore the study could be carried out only by drawing a representative sample and interviewing it several times. Such a panel study was made possible by a four-year grant, given in 1954 by the Ford Foundation to the Survey Research Center of the University of Michigan, for studies, directed by the author of this paper, on the origin and effects of economic attitudes.

FORMS OF ATTITUDE CHANGE

When at two successive points of time two different representative samples drawn from the same universe are asked the same question, or are given the same test, we obtain information on how the response has changed in that universe.

For instance, we may ask a representative sample of the American people about their leanings toward the Republican or the Democratic party; then, say, six months later we may ask another representative sample the same question. We will find either that no changes have taken place (or that the observed changes in the "marginals" are small and not significant—Case 1, Table 1) or that there were substantial changes from, say, Response B to Response A (Case 2).

Much additional information can be obtained from queries directed to two representative samples at two successive points of time. By asking several questions at both times, we may study the consistency or inconsistency of the observed changes. Or we may determine the demographic and economic characteristics (age, income, etc.) of those who gave the one or the other response. If, however, our major purpose is to analyze the changers—to find out why changes took place and what the consequences of

the changes were—we may use a different procedure, namely, to contact the identical people twice in succession. This procedure has disadvantages for the purpose of measuring the changes in distributions because of panel mortality and panel bias (6, 9), but it has the great advantage of yielding information on the turnover of individuals.

Changes in the over-all distribution, whether small or large, may come about either through few changes in the same direction or through many changes, some of which offset each other.² For instance, the "small" change in marginals illustrated in Case 1 (Table 1) from 50% to 55% A responses may come about by only 5% of the sample changing their response (minimum turnover) or by 95% doing so (maximum turnover), or by any proportion between these two extremes. Similarly, there are a great many possibilities as to how a change from 50% to

² This possibility has also been pointed out by Lazarsfeld (12, p. 517; 2, p. 243).

TABLE 1
MODELS OF MARGINAL CHANGE AND TURNOVER OF INDIVIDUALS

	Case 1 Measurement I II			Case 2 Measurement I II		
		50%	55%		50%	70%
Marginals		50	45		50	30
Response A		100%	100%		100%	100%
Response B						
Total						
Minimum Turnover	50A,A 0A,B 50IA	5B,A 45B,B 50IB	55IIA 45IIB 100	50A,A 0A,B 50IA	20B,A 30B,B 50IB	70IIA 30IIB 100
Intermediate Turnover	35 15 50	20 30 50	55 45 100	35 15 50	35 15 50	70 30 100
Maximum Turnover	5 45 50	50 0 50	55 45 100	20 30 50	50 0 50	70 30 100

Note.—I=first measurement; II=second measurement; A=frequency of Response A; B=frequency of Response B.

70% A responses may come about (Case 2). In this case the smallest possible number of changers is 20%, and the largest 80%.

This scheme applies to a great variety of situations. We may, for instance, think of experiments performed in the psychology of learning. Suppose that, at the time of the first test, the subjects do not know the answer to a question, and 50% say A and 50% say B. We then divide the subjects into two equated groups. Group 1, which we call the control group, is not taught anything between the first and second tests and happens to give 55%, rather than 50%, A responses the second time. Will this result come about by only 5% changing their response, or by many more shifting from A to B and also from B to A? The experimental group, Group 2, is given some instruction prior to the second test, to which we attribute the observed increase in the frequency of A responses from 50% to 70%. What proportion of individuals will have changed their response in this case?³

We may distinguish between the following four possibilities:

	MARGINAL CHANGE	TURNOVER OF INDIVIDUALS
Case 1a	small	close to the minimum
Case 1b	small	much larger than the minimum
Case 2a	large	close to the minimum
Case 2b	large	much larger than the minimum

Is it possible to characterize the situations in which each of the four cases is

³ Such an experiment might be performed, for example, by asking high school students who have no information about China: "Which is the largest city of China, Peiping, or Shanghai?" After recording the guesses, half of the students are shown a film about crowds of people on the Bund of Shanghai, while the other half does not hear about China prior to the second test. Before-after experiments, in which attitudes were measured both before and after the introduction of certain stimuli, are carried out frequently.

likely to occur? Or, given a small (large) marginal change, under which circumstances will we expect a "small" and under which circumstances a "large" turnover of individuals? Are there any principles which link either small or large marginal changes with either "small" or "large" turnover? In speaking of "small" turnover we mean a situation in which the number of changers is close to the minimum required to bring forth the given marginal change; and in speaking of "large" turnover, a situation in which the number of changers is much larger than the minimum. Therefore, here, as well as in our later discussion, the term turnover does not refer to the absolute number of changers.

Observed changes in the response are a function of true changes and of misclassifications. Misclassifications, either in the first or in the second measurement, may be due to interviewing errors, clerical errors, misunderstanding by the respondent, and intentionally or unintentionally incorrect replies (reporting errors). Effects of difficulties in the question which foster a misunderstanding, effects of guesses and hunches made because the respondent does not know the answer but is induced to reply, as well as effects of the mood of the respondent⁴ may be viewed in a certain sense as substantive changes. Yet they may be included under misclassifications so as to be excluded from other types of changes in response which permit us to study lasting effects of attitude change.

True change is then defined as a development (acquisition of experience, learning) which has taken place in the interval between the two measurements and affects the response. The develop-

⁴ These and related factors have been studied by Patricia Kendall (10).

ment may be personal, concerning just one individual in a group or sample; or it may be general, affecting many or all individuals in the group. The personal factors may cancel out for the group as a whole and thus influence the rate of turnover but not the marginal change, while the general changes may not cancel out and may even reinforce each other. The following example may serve to illustrate this difference. Suppose we ask a large sample of people: "Would you say that you are better off or worse off financially than six months ago?" in January and again in July. We may then find some people who had an accident or serious illness in the interval between the two contacts and who therefore shift from the answer "better off" to the answer "worse off." At the same time, we may find some people who, having recovered from an illness or accident, shift in the opposite direction. These are examples of true changes of a personal nature which may cancel out. If, however, there were general wage increases in March and April from which very many people benefitted, we would be confronted with a different kind of true change. Similarly, if business conditions deteriorated considerably from January to July, and newspapers, radio, and television brought extensive reports on wage cuts and profit declines, then we could expect a snowballing effect of the true change.

These considerations may help us to answer the question about the conditions under which the four cases are expected to occur. Some of the notions to be presented have received support through the studies of Kendall, while others will serve as hypotheses to be tested in our studies:

Case 1a: Small marginal change and

small turnover of individual responses. We expect to find this case when simple and easily understood questions call for well-established facts or attitudes and when there is no true change in the period between the two measurements. For instance, when the question concerns the education of adult respondents, we expect to obtain substantially the same marginal distributions in January and July and expect this to come about through relatively few individuals changing their response.

Case 1b: Small marginal change and large unnecessary turnover. Suppose again there is no true change but the question calls for guesses and hunches rather than for a well-established answer. Then a sizable proportion of people may shift from Guess A to Guess B, and another similar proportion from Guess B to Guess A.

We assume that this is not the only instance of Case 1b. True changes in both directions due to different personal experiences may likewise bring forth small marginal changes and large turnover, even if question and answer are clear and do not involve guesses.

Case 2a: Large marginal change and small turnover. The prototype of this case is, of course, effective learning by an entire group. For instance—we change the previously given example slightly—if prior to the second test our class is told that Peiping is the largest city of China, we shall expect many students to change their responses from Shanghai to Peiping, and hardly any from Peiping to Shanghai. In survey research we may expect a similar result under the following circumstances: A true change takes place which affects many people (is not a "private event"), is important, is easily understandable, and is reinforced through

social facilitation. For instance, assume that the cost of living went up radically in a given year, bringing hardship and considerable public discussion. Suppose, once before and once after the price increases, we ask: "Have prices of things you people buy gone up, remained the same, or gone down during the last year?" Then we would expect a considerable increase in the proportion of people answering "gone up" and hardly any frequency in the cell which represents shifts in the opposite direction (except for misclassifications).

Is Case 2a expected to occur under other less obvious circumstances? This is a crucial question for attitude research. Substantial aggregate shifts in intentions and expectations—such as an increase in the proportion of people expecting to vote for a candidate or expecting prices to go up—are often attributed some predictive value. Do such shifts arise through substantial turnover of individuals in both directions (Case 2b) or do they arise under circumstances which resemble the learning process, involve few counter-shifts, and may be classified as "social learning"?

Case 2b: Large marginal change and large unnecessary turnover. Unreliable measurements with true change in one direction or with a preponderance of misclassifications in one direction may bring about this case. To give just one somewhat artificial example: Suppose, once on a beautiful summer day and once on a dark winter day, we ask: "A week from today will the weather be cloudy or sunny?" Then it is possible that there will be many shifts in both directions, but the shifts from "sunny" in the summer test to "cloudy" in the winter test will predominate.

Alternatively, it is possible that Case 2b occurs even though people are not

guessing. Assume that both "public" and "private" events take place: for instance, that there is news of an improvement in the business situation and that some people have experienced income increases and some people income declines. Shall we then find a substantial marginal change coinciding with a large turnover of responses?

STATISTICAL MEASURES

Our first task is to construct measures for the two crucial variables, namely, marginal change and the rate of unnecessary turnover of individual responses. The measures should be applicable to instances in which there is no true change, as well as to those in which there is a true change in either or both directions. Furthermore, the measures should not be restricted to the deceptively simple dichotomous distributions which have been presented so far (in Table 1). We shall present measures for 3×3 distributions—attitudes measured twice on scales such as up-same-down or better-uncertain-worse—since most of our data are of this kind. Yet the measures can be extended to more elaborate data. The measures and this paper will be restricted to the analysis of turnover in two successive tests.

We introduce first a set of symbols. The marginals are denoted by small letters; those obtained in the first measurement are called p and in the second measurement, q . We have, of course, three cells for individuals giving unchanged or consistent responses (C), and six cells for individuals with changed responses; two of the latter represent two-step changes (denoted with the suffix 2), and four represent one-step changes.

		MEASUREMENT I			
MEASUREMENT II	C ₁	G _{1A}	G ₂	q_1	
	L _{1A}	C ₂	G _{1B}	q_2	
	L ₂	L _{1B}	C ₃	q_3	
	p_1	p_2	p_3		

There are two types of changes; we denote them with the letters G (gain) and L (loss).⁵ We call those changes G which bring forth the marginal change, and those changes L which detract from it. Viewing the marginal change as a current, gains represent swimming with the current; and losses, swimming against the current. By definition the gains are larger than losses (that is, the turnover tables will be written so as to conform with this definition).

By differentiating between two-step changes and one-step changes and arbitrarily assigning a double value to the former, we have

$$\begin{aligned} G &= 2G_2 + G_{1A} + G_{1B} \\ L &= 2L_2 + L_{1A} + L_{1B} \end{aligned}$$

Marginal change (M Ch) can be expressed either in terms of marginals or in terms of turnover cells:

$$M \text{ Ch} = G - L = (q_1 - q_3) - (p_1 - p_3) \text{ or } (2q_1 + q_2) - (2p_1 + p_2).$$

The expression $(100 + q_1 - q_3)$ or $(100 + p_1 - p_3)$ has been frequently used to compute an index of attitudes.⁶ Clearly the difference between two successive index values represents the aggregate or marginal change in attitudes.

The proportion of changers $(G + L)$ and the proportion of consistent people $(C = 100 - (G + L))$, disregarding the duplication of two-step changes) do not represent useful measures for the purpose of testing the relation of "unnecessary" turnover to marginal change. Obviously, the larger M Ch $(G - L)$, the larger is G (and therefore $G + L$). The crucial variable is L, as emphasized first by Lazarsfeld and Kendall (10). Yet we cannot rest satisfied with measuring turnover by L alone. L is negatively correlated with marginal change. What we are interested in is to find out where, in actual observations, L lies in the continuum between the smallest possible instance (minimum L) and the largest possible instance (maximum L). We locate the position

of observed L in the continuum between Min L and Max L:

$$\text{Min L} \xrightarrow{\text{L}} \text{Max L}$$

Our turnover measure, T, represents the relation of the distance between L and Min L to the distance between Max L and Min L. Since Min L is zero in all our distributions, we get:

$$T = \frac{L - \text{Min L}}{\text{Max L} - \text{Min L}} = \frac{L}{\text{Max L}}$$

This formula yields the value of 1 if L is equal to Max L, and a value of 0 if L is zero. In other words, the smaller the T, the smaller the proportion of changers beyond those necessary to bring forth the marginal change.

Given the marginals it is always possible to calculate Max L. In a simple, pragmatic way we calculate differently for two cases: those in which p_1 is larger and in which p_1 is smaller than q_3 . In the first case $2q_3$ represents the maximum possible value for the q_3 row; to this must be added the maximum possible value of L_{1A} , which is either q_2 or the difference between p_1 and q_3 , whichever is smaller. Therefore,

$$\begin{aligned} \text{Max L} &= 2q_3 + \min \{(p_1 - p_2) \text{ or } q_2\} \text{ if } p_1 \geq q_3 \\ \text{Max L} &= 2p_1 + \min \{(q_3 - p_1) \text{ or } p_2\} \text{ if } p_1 < q_3 \end{aligned}$$

The measures presented here do not represent a solution of the complex problem of developing independent indicators for extent of change in distributions (trend) and for the number of changers. For our purposes of relating unnecessary turnover (rather than number of changers) to changes in distributions, T appears to be a useful formula.

The formulas given above can be worked out for 3×3 distributions in which no distinction is made between one-step changes and two-step changes. In that case L is equal to the sum of the three Loss cells, while

$$\text{Max L} = \min [(p_1 + p_2) \text{ or } (q_2 + q_3) \text{ or } (p_1 + q_3) \text{ or } 50\%]$$

The data to be presented forthwith have also been calculated on this basis and indicate similar regularities. This method of calculating turnover involves, however, loss of information and is therefore inferior to the first method.

The statistical measures prepared to 3×3 distributions with duplication of two-step changes can be readily generalized for more complex distributions.

$$M \text{ Ch} = [(n-1)q_1 + (n-2)q_2 + \dots + q_{(n-1)}] - [(n-1)p_1 + (n-2)p_2 + \dots + p_{(n-1)}]$$

L can be calculated by multiplying the frequency of observed loss cells with the number of steps by which they are removed from the diagonal. Max

⁵ The letters G and L have been taken from information theory. The relation between information transmitted and information received has often been presented in a manner similar to our turnover tables. To be sure, in our case it is not possible to speak of "true information" (information transmitted and received), but only of consistent information or consistent attitudes. Yet what has often been called, in information theory, Gain and Loss—namely, information not transmitted but received, and information transmitted but not received—presents some similarities to our case of gaining or losing information over time.

⁶ See, for instance (9, p. 93) and the references given there to other diffusion indices.

L can be calculated by preparing a turnover diagram which places the maximum number of observations, consistent with the given distribution of the two sets of marginals, in the extreme lower left corner.

Our T measure is one of several possible measures. The literature contains, as far as the author is aware, two other measures. For the purpose of studying stability and instability of response, Lazarsfeld has developed a turnover index, x , which is applicable to dichotomous distributions only.⁷ The index is calculated from the equation, $L = x(R - x)$, where R is the sum of the marginal values of the row and the column in which L falls.

The Lazarsfeld index is derived from latent structure theory. The difference between latent true values and observed values is relevant, of course, for what we have called misclassifications. The index has been extended to circumstances in which there is a true change in one direction but, as expressly stated (II, p. 180), not to cases in which true changes may occur in both directions. The authors add that the assumption "that the true change takes place in one direction only is not easily defended when talking about attitudes." The Lazarsfeld index is not unrelated to our T measure. When the proportion of consistent cases is relatively stable over a set of items, x is positively correlated with T.

Ferber (I) set up a small consumer panel and asked the panel members identical questions several times at monthly intervals. He addressed himself to the following question: "To what extent have the replies given by the respondents to the same question remained the same over time periods of varying length?" (10, p. 257f.).⁸ The question calls for C, the proportion of people giving consistent responses, as a criterion. Since some C responses are expected by chance, Ferber uses $\frac{C^*}{\text{Exp } C}$ as his statistical measure; Exp C is

the consistency which might be expected by chance if the two responses were entirely independent, given the frequency of the marginals.

Although C is not a satisfactory measure for our purposes, we might think of following Ferber's lead and take $\frac{L}{\text{Exp } L}$ as the statistical meas-

ure of turnover. In making use of this measure, as in the case of using x , similar regularities were observed as with T. Yet we shall not use Exp L as our criterion because the relation of observed frequency of losses to chance frequency is of little interest to us. In our case the stability or change in answers is not random. It can be demonstrated that the second response is not independent from the first one and from the developments which have taken place in the period between the two measurements. The relative distance of observed L from maximum and minimum L concerns us much more than the relative distance of L from the chance situation. In addition, not only L but also Exp L is by necessity negatively correlated with M Ch. Sometimes, then, both L and Exp L are close to Min L, while at other times both are close to Max L; this crucial difference is not indicated by the measure $\frac{L}{\text{Exp } L}$.

In constructing the measure T, we have not taken misclassifications into account although clerical errors, interviewing errors, and reporting errors occur in all surveys and therefore, as set forth earlier, an observed change in attitudes is a function both of the true change and of misclassifications. Maccoby recently analyzed misclassifications resulting from unreliability of measurement in a perceptive article (15). Following her discussion, we may assume that misclassifications are random: an error is equally likely to occur for any response.⁹ In each test, then, misclassifications are proportional to the size of the response and will affect small cells differently from large cells. Since the L cells are smaller than the G cells (by definition), it can be demonstrated that the observed L is always an overestimate of the true loss (L without misclassifications).¹⁰ L is further away from Min L, or zero, than the true loss. If it is shown, as will be done in the next section, that in certain circumstances the observed L tends to be closer to Min L than in other circumstances, then it follows that such a statement would even be more true if it were possible to eliminate misclassifications. Because of the existence of misclassifications, the data used represent a strong test of the hypothesis which will be studied.

⁷ Presented in Kendall (10) and attributed to Lazarsfeld on page 180. Turnover, in Lazarsfeld's terminology, denotes the proportion of changers.

⁸ We shall be concerned here with Ferber's paper only insofar as it studies changes of response in two successive tests. The discussion of turnover in the course of several interviews has been postponed for a later article.

⁹ Randomness may rightfully be assumed for clerical errors. Reporting errors are known to be biased regarding certain financial variables (e.g., amounts saved). Regarding economic attitudes and expectations, there is no evidence of bias.

¹⁰ This is put by E. Maccoby as follows: "It will inevitably be true that a higher proportion of a minority group will shift" (15, p. 359f.).

QUANTITATIVE FINDINGS

Three different sets of circumstances may be distinguished under which turnover in two successive measurements can be studied. Some responses are expected to be stable over very long periods, as, for instance, sociocultural norms or tastes assumed to be innate or acquired in early childhood. Other responses are habitual and may not change during the short period between test and retest. Other responses again depend on changing circumstances and are expected to change in the short run (14, p. 1150; 10, p. 5; 7, Chap. 3).

While Kendall's study is directed toward "characteristics which are generally considered deep-rooted or basic" (10, p. 6), we are interested in attitudes making for short-run changes in behavior. Fluctuations in economic motives, attitudes, and expectations have been studied by the Survey Research Center over the past ten years in order to determine their influence on changes in spending or saving. Katona and Mueller have constructed an index of economic attitudes (9) and studied the relationship of changes in that index to changes in purchases of durable goods. Because of their association with changes in behavior, the attitudes which constitute that index, as well as a few other related attitudes, are particularly suitable for our present purposes.

Among the economic attitudes measured several times during the past few years, we shall find some whose distribution remained substantially the same in test and retest, and some whose distribution showed substantial changes. There also are several instances where the same attitude changed from Time Point 1 to Time Point 2 and remained stable be-

tween Time Point 2 and Time Point 3 (or vice versa).

Most of the data presented in this paper are taken from a panel study. A representative sample of the urban population of the United States was drawn. Lengthy interviews were conducted with members of that sample in June 1954, December 1954, June 1955, December 1955, and February 1957.¹¹ Fourteen attitudinal questions were asked both in June 1954 and June 1955 and yielded 14 turnover tables. In December 1955, 5, and in February 1957, 12 of the questions were asked again of the same respondents. Thus we have 31 turnover tables at our disposal which are constructed from questions asked of the same population and which differ in the following respects: (a) they reflect attitudes expressed at different times (in 1954, or 1955, etc.); (b) the time span between the two measurements varies (from 6 months to 20 months); (c) the attitudes differ. Some of the questions asked for evaluations of the recent past, while others were concerned with short-range or longer-range expectations. Questions were asked about people's attitudes toward their personal financial situation as well as toward national economic trends and market conditions.¹²

The 1954-57 panel study was not the first attempt of the Survey Research Center to interview identical samples twice. Parts of the samples of Surveys of Con-

¹¹ Several further publications based on studies with the same panel, made possible by the grant from the Ford Foundation, are in preparation. One of them will describe the sample in detail and will analyze panel mortality and panel bias. A substantial part of the questionnaires dealt with spending and saving.

¹² The reader will find the wording of most questions in the parts of this paper which present data derived from individual questions.

sumer Finances, conducted for the Federal Reserve Board, were interviewed both in 1948 and in 1949; and in 1953, part of the 1952 sample was reinterviewed. Five attitudinal questions on the first occasion and four on the second occasion were asked twice of the same samples at an interval of 12 months. Thus nine turnover tables derived from Surveys of Consumer Finances can be included in our set of data.

Most of our attitudinal questions are answered in the respondent's own words (open-ended questions), and the answers are taken down by the interviewer as nearly verbatim as possible. Coders classify the answers according to pre-established categories in the central office. The classifications include "don't know" and "not ascertained." Individuals classified in either of these categories in test or retest were omitted from the turnover tables. There remained individuals classified in either of three categories:

(a) up, better, or good; (b) same, pro-con, or uncertain and (c) down, worse, or bad.

Aggregative Change and Turnover of Individuals

A scatter diagram (Figure 1) indicates the relation of marginal change to turnover for the 40 repeated attitude measurements. At first sight the relationships between the two measures are not very clear. We may begin by discussing some of the points on the diagram in relation to the model which postulated the existence of four cases.

First we find some data which come close to Case 1a (small marginal change and small unnecessary turnover of individual responses). Take, for instance, the point on the diagram with an M Ch of 0.8 and a T of .39 and the point with an M Ch of 0.2 and a T of .40. Both points result from asking the same question and represent the turnover from June 1954

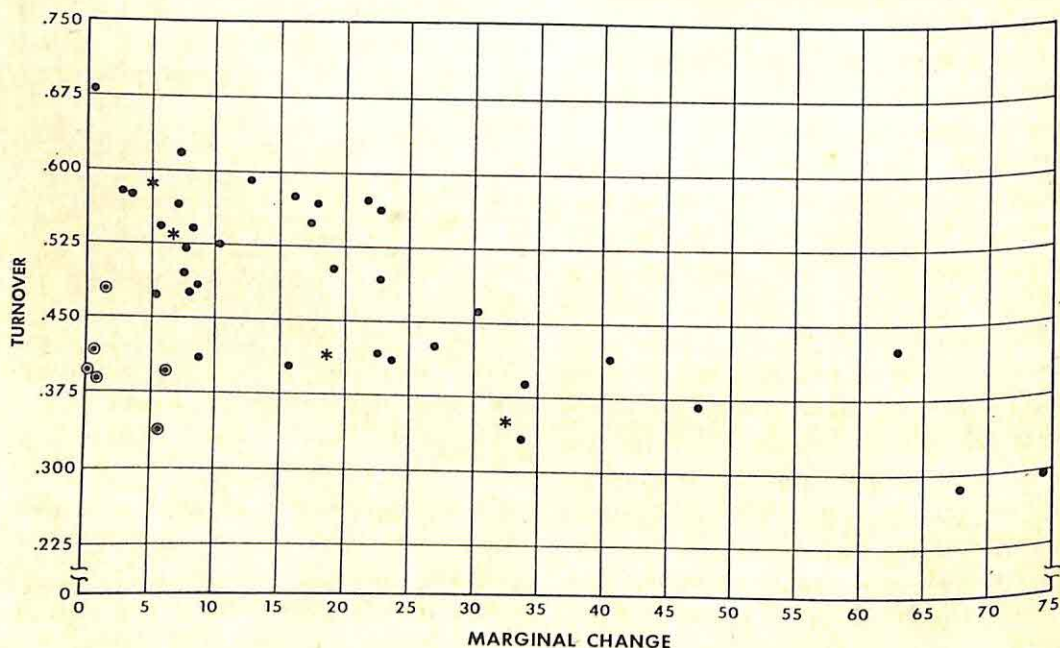


FIG. 1. Relation of Turnover Rates to Marginal Change. Each of the 40 attitudinal questions, asked twice in succession, is represented by a point. The asterisks represent clusters of attitudes.

to June 1955, and again from June 1955 to February 1957. The question read:

A few years from now, do you think you and your family will have a better position and income than you have now, or will you be in about the same situation, or even in a less satisfactory situation?

In all three surveys the distribution of the answers was practically the same. At the same time, relatively few individuals shifted. For instance, among the 827 people who gave definite answers to the question both in 1954 and in 1955, there were 6 individuals who shifted from better to worse and 4 individuals who shifted from worse to better. (These shifts may be due to misclassifications!) Although there were somewhat more people who shifted from same to better or from better to same, we find a small rate of turnover; L was 18.5 and much smaller than the highest possible L of 47.6. The data obtained in studying the turnover from June 1955 to February 1957 are similar.

Secondly, we find a number of instances resembling Case 1b (small marginal change and large turnover). As an example, we may refer to M Ch of 0.3 with a T value of .69. The question asked in June 1955 and December 1955 read:

Do you think that in the country as a whole during the next twelve months we'll have good times financially, or bad times, or what?

The two marginal distributions obtained could hardly be distinguished from each other. Nevertheless, among 757 people there were approximately 100 whose opinion was more pessimistic the second time than the first time and approximately 100 whose opinion was more optimistic the second time. L was 15.5 and relatively close to the maximum possible L of 22.6.

Case 2a (large marginal change and

small turnover) can be illustrated by the M Ch of 74.3 and the T of .30. The question read:

Would you say that at present business conditions in the country as a whole are better or worse than they were a year ago?

In June 1954, 17% said better and 53% said worse; in June 1955, 51% said better and 12% said worse. Among 818 respondents, only 8 shifted from better to worse. L was 8.7 and relatively distant from the maximum L of 29.3.

Similarly, we find a substantial marginal change in price expectations measured in 1948 and 1949 (67.8). Early in 1948 the largest proportion of consumers thought that prices of things they buy would go up during the next year, while early in 1949 the majority thought that prices would go down. L was 10.8, Max L 37.8, and T .29.

Instances of very large marginal changes coupled with large turnover (high T) are not found in our collection of turnover tables. But close to Case 2b comes, for instance, the M Ch of 21.8 with a T of .57. In answer to the question:

What do you expect prices of household items and clothing will do during the next year or so, stay where they are, go up, or go down?

many more said "down" in June 1954 than in June 1955; the shift occurred mainly from "down" to "same." At the same time, shifts in the opposite direction were not infrequent. L was 16.2 or closer to the Max L of 28.3 than to zero. (The data presented in Table 8, with M Ch of 17.9 and T of .57, may also be cited as resembling Case 2b.)

What determines the relationship between M Ch and T ? Is it the attitude itself? The assumption that responses to certain attitudinal questions will always be either of Type 1b or Type 2a can be

contradicted. A case in point is the question about expected business conditions in the country as a whole. It was given above as an example of Case 1b (small marginal change and large turnover) during the period from June 1955 to December 1955. The same question was also asked of the same sample in June 1954. The data for June 1954-June 1955 yielded the substantial marginal change of 40.6, many more people being optimistic regarding the economic outlook in June 1955 than in June 1954. The large marginal change from June 1954 to June 1955 was associated with a T value of .41, while the absence of a marginal change in the second half of 1955 ($M\ Ch = 0.3$) was associated with a T of .69, as reported above. Thus the same question asked from the same people once yielded a large marginal change and a small turnover, and once a small marginal change and a large turnover.

Similar findings were obtained with several other questions. For instance, the question:

Do you think now is a good time or a bad time to buy large household items such as furniture, refrigerator, TV, and things like that?

was also asked at the same three dates. From June 1954 to June 1955, $M\ Ch$ was 34.0; from June 1955 to December 1955, it was 2.5. In the first period, we obtained a T value of .39; in the second period, .58.

On the other hand, there are good reasons to believe that belonging to Case 1a (small marginal change and small turnover) is a function of the attitude studied. We find three questions in our collection regarding which one could have predicted a priori that they would yield small marginal changes and small turnover rates. Take first the following question:

How do you people feel about your present income; do you think it is about what you ought to be getting, or not?

This question, asked three times and yielding therefore two turnover tables, is easily understood and relates to relatively permanent or deep-seated notions which are not expected to change over short periods during which most people's income has not changed substantially. The same is true of the question quoted above which asks people whether their position and income will be better or worse a few years from now. Two turnover tables obtained through this question were cited as instances of Case 1a. Finally, we may include in the same category two turnover tables based on short-range personal expectations.

Do you think a year from now you people will be better off financially, or worse off, or just about the same as now?

The six observations relating to the three questions cited have been encircled on Figure 1.

That the three attitudes, yielding six turnover tables, represent a subset of our data which needs to be differentiated from the other data can be demonstrated by using a measure of consistency (C). The general assumption is that answers to certain attitudinal questions are more "personality determined" and less situationally determined, while it is the other way around with respect to answers to other questions (especially those which are intended to elicit subjective notions or shortrun attitudes dependent on changing circumstances). In the first instance we should expect less learning than in the second instance. In our turnover tables this should be apparent through a higher rate of consistency or repetitiousness in the first than in the second instance. A measure of repetiti-

TABLE 2
DATA ON MARGINAL CHANGE AND TURNOVER
A: Summary of All Turnover Tables

Marginal Change	Average of L	Average of Max L	Average of T	No. of Attitudes
Less than $3\frac{1}{2}$	23.8	48.2	.503	7
5 to $12\frac{1}{2}$	23.9	48.2	.509	14
16 to $30\frac{1}{2}$	17.8	37.5	.494	12
34 and larger	12.9	37.1	.358	7
				40

B: Summary of Turnover Tables Relating to Short-Run Attitudes

Marginal Change	Average of L	Average of Max L	Average of T	No. of Attitudes
Less than $3\frac{1}{2}$	23.8	39.8	.613	3
5 to $12\frac{1}{2}$	24.8	48.5	.518	12
16 to $30\frac{1}{2}$	17.8	37.5	.494	12
34 and larger	12.9	37.1	.358	7
				34

Note.—Rank correlation between M Ch and T = $-.689$ (Part B).

ousness has been developed¹³ and yields the highest values for the 6 observations encircled in Figure 1 and lower values for every one of the other 34 observations.

We shall now group our data according to the size of marginal change. Table 2A presents a summary of all 40 turnover tables at our disposal, and Table 2B the summary of 34 turnover tables. In Table 2B, the six observations which we deem to belong under Case 1a are omitted so

as to study the relation of small and large marginal changes to turnover rates for short-run attitudes. We find in Table 2B evidence for the following generalization: The larger the marginal change, the smaller the "unnecessary" turnover of individual responses. Subsets of our data, for instance the 1954-55 or the 1955-57 turnover tables considered separately, yield the same regularities.

In analyzing attitude changes it has been repeatedly argued that changes in answers to individual questions are much less reliable than changes derived from clusters of questions (9). Three such clusters have been computed from the material available in both June 1954 and June 1955. The first cluster contains six questions asking for evaluations of *present conditions*; the second, four questions regarding *one-year expectations*; and the third, three questions regarding *five-year expectations*. Finally, we have at our disposal the turnover in the *index of consumer attitudes*, for which time series have been published recently (9); this index consists of six questions, some of which fall into each of the three clusters. Table 3 contains the data on the turnover of our four group measures. We again find that, the larger the marginal change, the smaller is T.

Conclusions

We may now formulate the following generalization applicable to subjective notions and expectations which do not represent deep-seated and enduring convictions:

If two successive measurements yield similar aggregate distributions, then it is likely that many individuals have changed their attitudes in one, and many others in another, direction. If in the aggregate there is a substantial change in

¹³ The statistical measure of repetitiousness or variability will be presented in forthcoming papers dealing with change in attitudes over several surveys (a problem mentioned in Footnote 7) and with changes in nonverbal behavior (spending, saving) over several years. The measure derives from a computation of coefficients of intraclass correlations. It reaches its maximum value when the frequency of identical answers (or of identical behavior) is maximized, a zero value when the actual observations correspond to what would be expected if the consecutive observations were independent of each other, and a minimum value when the frequency of identical answers (or behavior) is minimized. The measure has been developed by Leslie Kish of the Survey Research Center.

attitudes, then it is likely that the changes will be predominantly in one direction; the population then tends to fall into two major groups: those who shifted in one direction, and those who did not change at all.

These findings were not unexpected; they are in accord with hypotheses formulated before embarking on the studies. Although the findings do not prove the correctness of the hypotheses, it may be mentioned that the hypotheses were derived from sociopsychological assumptions about contagion and social learning, as well as from earlier findings obtained without the use of the panel technique. When in earlier years consumer attitudes showed large changes in the aggregate (as, for instance, in 1950-51), all subgroups of the population showed similar changes. When, however, measurements with two successive samples indicated substantially unchanged distributions of attitudes (for instance, in 1952), some occupational, regional, or income groups showed a shift in one direction and some other groups in the opposite direction.

From the point of view of utilizing survey research, our findings indicate first that small changes in the distribution of subjective notions and expectations must be viewed with caution. Substantially unchanged distributions, obtained at two successive occasions, do not imply that very few people changed their opinions. On the other hand, substantial shifts in aggregative distributions may be viewed with much less skepticism. To be sure, it is always valuable to make certain through panel data that most individual shifts were in the same direction. But even without recourse to panel data, such an assumption appears to be justified in most cases.

Our generalization may be better un-

derstood if we contrast it with the discussions of Lazarsfeld and Kendall. In the theoretical scheme of Lazarsfeld, "turnover," that is, the proportion of changers, is an index of uncertainty or instability:

If the turnover is large, it indicates that the opinion or behavior is unstable. We know that people feel uncertain and that propaganda may be effective, or that clarification and education are required (13, p. 232).

Similarly, Kendall identifies "turnover" with instability of response due to vacillation by respondents as shown, for instance, in the following passage:

Instability of response (turnover): Why is it that, under certain conditions, respondents vacillate in their answers to repeated questions? (10, p. 29).

No doubt, these authors correctly describe one type of change in response. It does happen that respondents vacillate or that the response is unstable because clarification and education are required. But our findings indicate that responses may also change under different circumstances. Rather than being the result of uncertainty, change may be due to acquisition of experience or learning which took place on the part of very many people at about the same time. If a high proportion of changers is associated with "uncertainty" (or with contradictory developments of a personal nature), the data will resemble Case 1b in our model (relatively high T value); if a high proportion of changers is associated with "social learning," the data will resemble Cas 2a (relatively low T value).¹⁴

¹⁴ Turnover due to true changes rather than to vacillations has been considered by Lazarsfeld and his associates in connection with before-after experiments. In discussing, for instance, the showing of a film on antisemitism to people whose level of antisemitism was measured both before and after the showing of the film, Glock speaks of "the effect of a stimulus in producing

On the Origin of Changes in Attitudes

When in two successive surveys some people express the same attitude and some change their attitudes, we have two methods at our disposal for explaining the difference between the two groups. We may determine whether the two groups—those with stable attitudes and those with changing attitudes—differed in other respects in the *first* survey, with the aim of finding out whether any such initial differences may account for what happened to their attitudes later. Alternatively, we may study data obtained in the *second* survey to find out whether any developments occurred in the period between the two surveys which would differentiate between the two groups. The two kinds of studies will be called Approach I and Approach II, respectively.

These two approaches to the study of the origin of stability or change in attitudes are related to the distinction previously made between misclassifications, in the broadest sense, and acquisition of experience. Naturally, it is only by inserting appropriate questions in the second survey that we can study whether new experiences, either of a personal or a general nature, have been acquired in the interval between test and retest. At the other extreme, observed changes in attitudes may be due to incorrect recording in the one or the other survey. More important, we may think of the answers to a question as falling on a continuum: Some people who say "better" the first time may have very different attitudes

from those who say "same," but it may also happen that the true attitudes were relatively close and were distinguished somewhat arbitrarily in our crude system of measurement. In the latter case a recorded shift from "better" to "same" may not represent a significant change.¹⁵ Thus it may happen that some people who gave Answer A did not really belong to the A group and may be expected to shift to Answer B in the light of characteristics or other attitudes measured in the first survey.¹⁶

The two approaches are not alternatives. Holding an attitude with little conviction may facilitate the acquisition of experiences which result in changing that attitude. Thus, we should expect to find circumstances in which both Approach I and Approach II contribute to an "explanation." Yet, interestingly enough, at least one case will be found in which Approach I will not result in an explanation and Approach II will.

Findings: Origin of Changes in Two Short-Run Expectations

Detailed studies were carried out regarding two of the turnover tables summarized before. Table 4 presents the basic data. Both questions were asked of the same sample, once in June 1954

¹⁵ Eleanor Maccoby calls attention to this possibility and illustrates it with a graph (15, p. 361).

¹⁶ Procedures similar to what has been called here Approach I have been used in analyzing shifts by election panels. For instance, those whose intended vote shifted, say, from the Republican to the Democratic candidate were studied regarding their "class interest" (2, p. 247). It was found that "Among those people who planned to vote for Dewey in August, those whose class interest was in conflict with their vote intention were much more likely to abandon Dewey than those where such conflict did not arise." Such initial differences were sometimes called cross-pressures (12, p. 512).

change" in attitudes (2, p. 243f.). Under such conditions the findings about the rate of turnover were often similar to our findings on attitude changes resulting without experimental stimuli. Sometimes, however, even though the experimental stimulus brought forth a substantial marginal change, reverse changes were also observed and were called the "boomerang effect."

TABLE 3
MARGINAL CHANGE AND TURNOVER FOR CLUSTERS OF ATTITUDES

	M Ch	M Ch ^a	L	Max L	T
5-year expectations (6×6 table)	12.7	5.0	354	602	.588
1-year expectations (12×12 table)	36.1	6.6	534	996	.536
Present conditions (12×12 table)	179.4	32.6	325	929	.350
Index of consumer attitudes (12×12 table)	102.0	18.5	356	878	.416

^a M Ch recalculated so as to make it comparable to 3×3 tables. The data presented on the size of marginal change in the first column are not comparable to those presented in Table 2. In a 12×12 table the maximum number of steps with which changed responses are multiplied is 11; in a 3×3 table it is 2. If, then, we divide M Ch as presented in the first column of the table by 11/2 (or in one instance by 5/2), we obtain an M Ch of 5.0 for five-year expectations, of 6.6 for one-year expectations, of 18.5 for the index, and of 32.6 for present conditions. These data with their respective T values have been added to Figure 1 in the form of asterisks.

and once in June 1955. One of the questions:

Do you think that a year from now you people will be better off financially, or worse off, or just about the same as now?

resulted in a very small marginal change. The other question:

Do you think that in the country as a whole during the next twelve months we will have good times financially, or bad times, or what?

resulted in a substantial marginal change. These two questions were chosen so that the origin of changes in both personal financial and business expectations could be studied. The turnover rates obtained in the two instances are similar; neither is substantial.

Table 5 contains the data obtained by applying the two approaches to the first question. Groups OO and OM, as well as Groups MM and MO, expressed the same attitudes in June 1954. Nevertheless, there are substantial differences between the pairs of groups in income, in age, and in education. Those who became more pessimistic (OM) have lower average incomes, are older, and have less education than those who remained optimistic (OO). Similarly, those who became optimistic (MO) have somewhat

TABLE 4
BASIC DATA FOR STUDIES OF ORIGIN
OF CHANGES IN EXPECTATIONS

A. Expected Personal Financial Situation during Next Year June 1954					
		Better (O)	Same (M)	Worse (P)	
June 1955	Better (O)	20.5	14.7	1.3	36.5
	Same (M)	13.3	40.6	3.6	57.5
	Worse (P)	1.2	3.9	0.9	6.0
		35.0	59.2	5.8	100.0%

Note.—M Ch = 1.3; T = .48; N = 850. Group OO, 20.5%, in first measurement to be compared with Group OM (including OP) of 14.5%. Group MM, 40.6%, in first measurement to be compared with Group MO of 14.7%.

B. Expected National Business Conditions during Next Year June 1954					
		Good (O)	Pro- con (M)	Bad (P)	
June 1955	Good (O)	46.8	20.0	13.5	80.3
	Pro-con (M)	6.0	7.0	3.2	16.2
	Bad (P)	1.5	0.6	1.4	3.5
		54.3	27.6	18.1	100.0%

Note.—M Ch = 40.6; T = .41; N = 844. Group OO, 46.8%, in first measurement to be compared with Group OM (including OP) of 7.5%. Group MM, 8.4% (including PP), in first measurement to be compared with Group MO of 23.2% (including PM).

higher income, are younger, and have more education than those who remained pessimistic (MM). Personal financial developments prior to the first test likewise contribute to an explanation of attitude change. In the OO group we find more who were better off than in the OM group, and in the MO group more than in the MM group. Finally, the evaluation of prevailing conditions in the first survey was more optimistic in the OO than in the OM group. Thus we conclude that data from the first survey help account for the difference between stable and changing expectations to a considerable extent.

Additional explanations are provided by the second approach. The proportion of Group OO whose income rose and whose evaluation of their personal financial situation improved is much higher than the proportion of Group OM. Between MM and MO we likewise find differences in the expected direction, but they seem to be less pronounced. People's opinions about the conditions in the industry in which they work also show differences which indicate that what happened to the two pairs of groups is not unrelated to stability or change in their personal expectations. It must be stressed that we explain group differences. As seen in Table 5 there are individuals in Group OM (the group which became more pessimistic) who said they were better off and whose income increased. Accounting for changes in attitudes by individuals is a task not undertaken here; it would require a different kind of interviewing.

The last part of Table 5 contains data related to acquisition of information about business conditions in the nation. The members of the panel were asked several questions in order to determine

whether or not they had "correct information" about what had happened to the economy in the year between the two measurements.¹⁷ It appears that Group OO was somewhat better informed than Group OM. There are hardly any such differences between Groups MM and MO.

We conclude, therefore, that the stability or instability of personal financial expectations is well accounted for, both through initial differences between the stable and unstable groups and through changes in the personal finances of the groups in the year between the surveys. This statement is especially true of the differences between Groups OO and OM.

Turning to an analysis of stability or change in the one-year economic outlook (Table 6), we find consistent differences between Groups OO and OM when we apply our first approach. The small deviant group which became more pessimistic has lower incomes, is somewhat older, has less education, and has evaluated the prevailing conditions less favorably than the group which remained optimistic. But the differences between Groups MO and MM in all these respects are extremely small or nonexistent. The same is true of changes in personal financial conditions prior to the first test (not shown in the table). Thus improvement in the general economic outlook, which in this case brought about a substantial marginal change, cannot be accounted for by initial differences between the changers and nonchangers.

Approach II reveals that, between the two surveys, changes occurred in income and in people's evaluation of their personal financial situation. These develop-

¹⁷ See the Appendix for a description of the technique used.

TABLE 5
PERSONAL FINANCIAL EXPECTATIONS
June 1954-June 1955

APPROACH I	OO	OM	MO	MM
1954 Median income	\$6,570	\$5,500	\$5,300	\$4,870
Median age	39	44	42	51
Education:				
Grade school	20%	33%	24%	48%
High school	48%	51%	53%	35%
College	32%	16%	23%	17%
Personal finances prior to first test:				
Better off	56%	39%	32%	21%
Worse off	20%	21%	26%	20%
Difference	+36	+18	+6	+1
Index of "Evaluation of Present Conditions" in first test:				
Average	7.2	5.7	6.1	5.9

APPROACH II

Personal finances in year between two tests:				
Better off	67%	32%	36%	18%
Worse off	8%	21%	18%	18%
Difference	+59	+10	+18	0
Income change in year between two tests:				
Making more	55%	20%	40%	22%
Making less	9%	21%	19%	12%
Difference	+46	+8	+21	+10
Conditions in respondent's industry in year between two tests:				
Better	54%	30%	49%	27%
Worse	7%	17%	13%	18%
Difference	+47	+13	+36	+9
Information received on economic developments during year between two tests:				
Hardly any correct information	11%	32%	26%	32%
Some correct information	60%	52%	54%	50%
Much correct information	29%	16%	20%	18%
Mean number of questions answered correctly	4.45	3.44	3.89	3.48

TABLE 6
ONE-YEAR ECONOMIC OUTLOOK
June 1954-June 1955

APPROACH I	OO	OM	MO	MM
1954 Median income	\$6,250	\$5,000	\$4,600	\$4,400
Median age	46	49	46	48
Education:				
Grade school	29%	39%	42%	44%
High school	42%	45%	42%	42%
College	29%	16%	16%	14%
Index of "Evaluation of Present Conditions" in first test:				
Average	7.1	5.7	5.8	5.5

APPROACH II

Personal finances in year between two tests:				
Better off	39%	25%	30%	23%
Worse off	14%	25%	15%	28%
Difference	+25	0	+15	-5
Income change in year between two tests:				
Making more	37%	22%	32%	23%
Making less	10%	25%	12%	33%
Difference	+27	-3	+20	-10
Had conversation on economic trends in year between two tests:				
Optimistic	15%	6%	8%	6%
Pessimistic	6%	17%	12%	16%
Difference	+9	-11	-4	-10
Information received on economic developments during year between tests:				
Hardly any correct information	14%	40%	27%	63%
Some correct information	56%	55%	55%	33%
Much correct information	30%	5%	18%	4%
Mean number of questions answered correctly	4.39	3.04	3.64	2.15

ments seem to contribute toward the explanation of the attitudinal change. More pronounced still are the differences in the extent of correct information received on economic developments that have taken place during the year between the two tests. In Group OO we find many more people who answered substantially all questions correctly than people who answered hardly any questions correctly; in Group OM the latter are very frequent and the former practically nonexistent. Similarly, very sizable differences exist between the MO and the MM groups, Group MO being much better informed. The differences in information between these groups are much more pronounced than those obtained in Table 5. More important still, one may discount the differences in information in Table 5 by assuming that they reflect primarily differences in education. In Table 6, we find hardly any differences in education between Groups MM and MO, and yet the differences in economic information are substantial.

Table 6 contains some further data which point in the same direction. In answer to the question:

During the last few months have you discussed with other people whether business conditions are getting better or worse?

substantially the same proportion of people in each of our four groups answered in the affirmative. But, in replies to a follow-up question which inquired about the content of the conversations, we find that some groups had more conversations of an optimistic nature and others of a pessimistic nature. These differences are again in the expected direction.

Since correct information about developments in 1954-55 referred, without exception, to favorable developments, we

may restate the findings as follows: Two of the groups, OO and MO, acquired information about favorable developments in the economy prior to the second survey. Such learning appears to be reinforced through social facilitation and represents a current which is hard to swim against. The acquisition of information corresponds with a change in attitudes on the part of Group MO. It probably also corresponds with a change in attitudes on the part of members of Group OO; here, however, our crude measurements were unable to detect a change in attitudes because those who gave the most optimistic recorded answer in the first survey ("good") could not improve their attitudes further.¹⁸

Thus somewhat different explanations emerge for the frequent improvement in attitudes toward national business conditions on the one hand (Group MO in Table 6), and the much less frequent improvement in people's attitudes toward their personal financial situation (Group MO in Table 5) on the other hand. Since we are dealing with changes in two attitudes among the same people at the same time, a further step in the analysis is possible: a comparison of the changes in the two attitudes. Of particular interest are people whose personal financial expectations deteriorated from June 1954 to June 1955 (the L group in Table 4A consisting of 18.5% of the sample). How did the business expectations of these

¹⁸ What has been said about Group MO also applies to Group PO (13.5%). Data on PO are not shown in Table 6 and are not conclusive because the corresponding PP Group is too small for comparisons. Yet it may be added that, according to Approach I: PO does not differ from MM, income gains are less frequent among PO than among MO, but PO has a substantial amount of correct information.

people change between the same two dates? We find that: (a) 8.0% fell in the cell OO in Table 4B—remained optimistic regarding business conditions; (b) 6.5% fell in cells MO, PO, or PM—became more optimistic; (c) 2.0% fell in cells MM and PP—remained relatively pessimistic; (d) 2.0% fell in cells OM, OP or MP—became more pessimistic.

Among people classified under *d* and possibly under *c* we may discern correspondence between the two changes, or even an influence of personal financial expectations on business expectations. Classifications *a* and *b* contain, however, a much larger proportion of the people whose personal financial attitudes deteriorated. The great majority of this group maintained optimistic business opinions or even strengthened them. If information on general economic developments is relevant for the improvement in business expectations, we must conclude that a pessimistic turn in opinions about personal finances lost in competition with good business news. The economic information constituted a current which was difficult to resist.

Also of interest is a second group: those whose personal financial expectations remained unchanged on a nonoptimistic level (Groups MM and PP in Table 4A, 41% of the sample). In checking how the national business outlook of these people changed at the same time, we find that 17.5% were optimistic both times, 16.5% became more optimistic, and only 7.5% remained pessimistic or became pessimistic. Again, among most of these people, business attitudes developed in a manner different from personal financial attitudes.

The same conclusion may be reached by studying personal financial experiences rather than changes in personal

financial expectations. As seen in Table 6, some members of the sample reported in June 1955 that they were personally worse off at that time than a year earlier. Yet almost three-fourths of those who made such reports fell into cells OO, MO, PO, or PM in Table 4B. Thus the business expectations of the great majority of people who said they were worse off financially were unaffected by their personal experiences.

Findings: Origin of Changes in Attitude Index

The factors contributing to stability and change in attitudes were also studied regarding the cluster of attitudes represented in the index of consumer attitudes. Among the six questions included in the index are the two analyzed in Tables 5 and 6. Each optimistic answer was given a value of 2, the middle position a value of 1, and each pessimistic answer a value of 0. Therefore, we have a 12-point scale with 12 being the most optimistic value. The turnover from the first to the second survey has been classified in eight groups.

Three of these groups are made up of people who have been consistent. Consistency is defined as having the identical scale value in both surveys (June 1954 and June 1955) or being one step removed. In other words, say, 8 in the first survey and either 7, 8, or 9 in the second survey is called consistent. We define Group OO as having scale values of 10, 11, or 12 in the first survey; MM as having scale values of 7, 8, or 9 in the first survey; PP as having scale values of 6 or less in the first survey; all three groups changed by one point or less.

Then we have three groups which show gains, that is, became more optimistic in the second than in the first survey. Group PM was fully comparable with

Group PP in the first survey (scale values of 6 or less) and became more optimistic in the second survey (increase in scale value by 2, 3, or 4). Group PO differs from PM in that it became much more optimistic in the second survey (increase by 5 or more). Group MO shows gains from a middle position. There are no gainers among those who were optimistic in the first survey.

We have two loss groups. Group OM had a scale value of 10, 11, or 12 in the first survey and became more pessimistic (decreased by two or more steps). The small Group MP consists of people with scale values of 7, 8, or 9 in the first survey who were more pessimistic in the second survey than in the first.

Approach I (Table 7) contributes to an explanation of stability or change. The differences in income are especially large among those who were either O or M in the first survey, and the differences in age among those who were P in the first survey. Substantial educational differences are found among all groups. Yet the differences in the evaluation of "present conditions" are small or, in the case of the P groups, contrary to expectations.

Turning to data obtained in the second survey (Approach II), we find that income changes during the year between the two surveys were substantial and contribute to an explanation without exception. The same is true of the appraisal of changes in the condition of the industry to which respondents belong. Differences in the nature of conversations on economic trends likewise indicate that loss groups had more pessimistic conversations than consistent groups or gain groups. Finally, looking at the data on information received, we find very substantial differences among the three groups which were M in the first survey.

We conclude then, in the light of data presented in Table 7, that initial differences among the groups, personal developments in the year between the two tests, as well as information obtained on national economic developments may account for stability or change in the index of consumer attitudes. Some of the differences, such as becoming or not becoming more optimistic on the part of originally pessimistic people, seem to hinge primarily on income changes and related personal developments. Regarding some other changes, interaction of various influences must be given a greater role. This is not surprising because the index is constructed both from questions relating to personal finances and from questions relating to expected economic conditions.

In analyzing the findings of this section, it must be recalled that the origin of stability or change in attitudes was studied in a period in which economic conditions in the country and the financial situation of most people improved considerably. Under these conditions deterioration of attitudes by a few people was found to be associated with having little information about economic developments, with personal events of an adverse character, and with "misclassifications" (the initial relatively favorable attitudes being not too trustworthy). Turning to an attempt to understand the major change which took place during the year—the improvement in attitudes by a substantial number of people—we find: Regarding personal financial attitudes, people who become more optimistic differed from people who did not by having more favorable personal developments (as well as by having characteristics and other initial attitudes which make them more predisposed to a favor-

TABLE 7
INDEX OF ECONOMIC ATTITUDES (6 QUESTIONS)
June 1954-June 1955

	OO	OM	MO	MM	MP	PO	PM	PP
Number of cases	136	64	140	160	28	55	87	30
Per cent of sample	19.9	9.2	19.8	22.9	3.9	7.9	12.2	4.2
APPROACH I								
Median income	\$7,750	\$5,950	\$5,900	\$5,220	\$4,600	\$4,450	\$4,450	\$3,970
Median age	49	46	43	47	50	40	49	53
Education:								
Grade school	21%	32%	18%	40%	50%	26%	47%	53%
High school	45%	44%	53%	40%	26%	65%	38%	44%
College	34%	24%	29%	20%	14%	9%	15%	3%
Index of "Evaluation of Present Conditions" in first test:								
Average	8.5	7.9	6.4	6.5	6.0	2.9	4.0	3.7
APPROACH II								
Income change in year between two tests:								
Making more	46%	29%	53%	27%	11%	44%	22%	13%
Making less	6%	17%	6%	13%	46%	16%	16%	43%
Difference	+40	+12	+47	+14	-35	+28	+6	-30
Conditions in respondent's industry in year between two tests:								
Better	46%	35%	43%	33%	11%	32%	22%	10%
Worse	8%	21%	4%	12%	25%	14%	13%	17%
Difference	+38	+14	+39	+21	-14	+18	+9	-7
Conversation on economic trends in year between two tests:								
Optimistic	23%	15%	10%	10%	0	11%	10%	3%
Pessimistic	6%	9%	11%	7%	29%	5%	10%	17%
Difference	+17	+6	-1	+3	-29	+6	0	-14
Information received on economic developments during year between two tests:								
Hardly any correct information	10%	14%	13%	24%	64%	35%	42%	40%
Some correct information	53%	62%	63%	55%	29%	58%	39%	53%
Much correct information	37%	24%	24%	21%	7%	7%	19%	7%
Mean number of questions answered correctly	4.68	4.29	4.35	3.82	2.13	3.32	3.29	2.77

able change). Regarding national economic attitudes, the changers are characterized by more frequent favorable personal experiences and by the acquisition of much more extensive favorable economic information. The same two considerations account for those who show an improvement in their index score.

Attitude Change in a Period in Which No Consistent Information Was Received

Up to now we have studied the origin

of changes in attitudes between June 1954 and June 1955, that is, in a period when the country's economic situation improved considerably. Further analysis is possible for the year 1956 with the same panel of respondents (based on surveys carried out in December 1955 and March 1957). In this second period economic activity on the whole remained at a high level, without showing significant changes; some sectors of the economy softened while others continued to improve. That the two periods would show

such differences was, of course, not known in advance. Yet the differences present us with an opportunity to compare two situations.

It may be useful to present a few data about economic developments during these periods. The data given refer to the quarters of the year in which the surveys were conducted. First, during the 1954-55 period: the index of industrial production rose from 124 to 138; an increase in personal incomes from 286 to 304 billions was used by consumers to step up their purchases of durable goods at a rate almost unprecedented in American economic history—from 28.8 to 35.3 billion dollars (these are seasonally adjusted annual rates for the respective quarters); and prices remained stable. Next, during the period from December 1955 to March 1957: industrial production did not change much, but prices advanced significantly (the consumer price index rose from 114.9 to 118.6); the rate of increase in personal incomes was as large as in the previous period in current dollars (increase from 315 to 336 billions), but not in real terms; consumers used their higher incomes to spend more on food and services and the like, while purchases of new automobiles were substantially smaller in 1956 than in 1955. Total consumer expenditures, considered in real terms, were at approximately the same level in the third quarter of 1955 and the first quarter of 1957.

In the first period, consumer attitudes improved considerably. Each of several questions about prospective economic trends in the country yielded large marginal changes (increased frequency of favorable answers). In the second period, most attitudinal questions showed a slight turn toward more pessimistic answers. Really substantial changes in the

distribution of answers occurred only when consumers were queried about recent past trends in prices and economic activity. Very many people gave correct, and therefore different, answers in 1957 and in 1955 to such "information questions." For instance, in June 1955, 51% of consumers said that times were better than a year earlier, and 11% said that they were worse; in March 1957, 23% said that times were better, and 29% said that they were worse (the remainder said that there was no change).

Among questions on short-range expectations which are considered in this paper, the one referring to expected economic developments during the next 12 months showed the greatest change in 1956. This is the question which is analyzed in Table 6 for the first period and which we shall now study from the point of view of the origin of changes during the second period. The basic data, presented for the first period in Table 4B, are shown for the second period in Table 8. They indicate that the marginal change was brought about by a fairly substantial turnover of individual attitudes. (The T rate is closer to the maxi-

TABLE 8
EXPECTED NATIONAL BUSINESS CONDI-
TIONS DURING NEXT YEAR
December 1955

March 1957	Good (O)	Pro-con (M)	Bad (P)	
Good (O)	59.0	7.2	2.3	68.5
Pro-con (M)	15.3	4.6	1.6	21.5
Bad (P)	7.3	1.4	1.3	10.0
	81.6	13.2	5.2	100.0%

Note.—M Ch = 17.9; T = .57; N = 717. Group OO, 59.0%, in first measurement to be compared with Group OM (including OP) of 22.6%. Group MM, 5.9% (including PP), in first measurement to be compared with Group MO of 11.1% (including PM and PO).

TABLE 9
ONE-YEAR ECONOMIC OUTLOOK
December 1955-March 1957

	OO	OM	MO	MM
1956 median income	\$6,000	\$5,300	\$4,100	\$4,750
Personal finances in year prior to second test:				
Better off	42%	27%	33%	23%
Worse off	12%	26%	24%	30%
Difference	+30%	+1%	+9%	-7%
Information received on economic trends during year before second test:				
Hardly any correct information	14%	27%	33%	35%
Some correct information	45	49	47	49
Much correct information	41	24	20	16
Mean number of questions answered correctly	4.07	3.48	3.25	3.16

mum possible T rate than to zero.) In looking for an explanation of these changes, only a limited number of data were found to be relevant; therefore we shall restrict our presentation to them.

We shall look first at the group of people whose attitudes improved (Group MO). In the first period (Table 6), this was a large group which was responsible for the over-all improvement; in the second period (Table 9), this was a small group whose attitudes partly offset more frequent changes in the opposite direction. We find in Table 9 that the incomes of Group MO are slightly lower than of Group MM; therefore the attitude change cannot be explained by differences in income. Correct information received on economic trends likewise does not contribute to an explanation of changes in attitudes. In the first period, this was the crucial indicator: Group MO knew much more about the favorable developments in the economy than Group MM. In the second period, there is no such difference. There remain some differences in the evaluation of personal financial developments. Although in both Group MO and MM there were many people who felt better off, as well as many people who felt worse off, there are differences between

the two groups in the expected direction.

Turning now to those whose attitudes became more pessimistic (Group OM), we may recall that here again, as in Table 6, several significant explanatory variables could be discerned, including large differences in the amount of correct information received between Groups OO and OM. In the second period (Table 9), we find some differences, although most of them are not quite as pronounced as in the first period. Group OM is not quite as well informed as Group OO, is more frequently worse off, and has somewhat lower income.

While in the first period correct information represented, without exception, favorable information, in the second period news was conflicting and a well-informed person acquired information about both favorable and unfavorable economic developments (see Appendix). Therefore, one may think of comparing people who knew of favorable trends with people who knew of unfavorable trends, rather than of contrasting well-informed and badly informed people. This type of analysis was carried out with two questions, the answers to which showed the most pronounced split among the respondents. About the same proportion of the panel said (a) that unemploy-

ment increased in the year prior to March 1957 as said that it did not increase, and (b) that fewer cars were sold in 1956 than in 1955 as said that this did not happen. The answers on unemployment show some relation to attitudes: In Group OM we find a much larger proportion of those who thought that unemployment increased than in Group OO. The answers on automobile sales do not show much relation: the proportion of people who knew that car sales were lower in 1956 than in 1955 was similar in the different groups. Possibly the interpretation of this information—the notion of how the decline in automobile sales would affect the future course of the economy—differed among members of the groups. Alternatively, it is possible that information received in the second period should not be viewed as an explanatory variable for observed changes in attitudes. In any case, in the second period we cannot speak of a “current of information” which uniformly influenced the sentiment of broad groups of people.

We learn from this analysis that under certain conditions—such as those which prevailed in the first period (1954-55)—information is salient, uniform, and influences people's opinions and attitudes in the same direction. Under other conditions, when conflicting news about slight favorable and slight unfavorable trends are transmitted—as in the second period (1956)—uniform learning does not take place and some people's attitudes change in the one, and some people's attitudes change in the other, direction. Probably the differences between the two situations depends both on the extent of developments (large or small changes in the economy) and on the uniformity of changes (whether or not the reported changes are all in one direction). In the

first situation, which is most relevant for business cycle analysis, it was possible to establish positive relations between information received and attitude changes; while in the second situation, it was not possible to do so.

Some Effects of Attitude Change

Attitudes are predispositions to action. A major purpose of studying expressed opinions and expectations is to relate them to nonverbal behavior. The function and predictive value of economic attitudes has been analyzed on the aggregative level by relating changes in the distribution of attitudes (in samples representative of all consumers) to durable goods purchases of the American people (see especially 9). Tests with individuals have been carried out by studying the purchases of people who are or have become optimistic in comparison to the purchases of people who are or have become pessimistic (16). A few such comparisons, relating to groups singled out for study in the preceding section, will be presented here. As an introduction, it must be said again (see 6) that individual tests are expected to show a weaker relation of attitudes to purchases than are aggregative tests. One of the reasons for this expectation is that individual spending behavior is influenced by a variety of factors which may cancel out in the aggregate. For instance, an optimistic person may not make any purchases in a given period because he has bought several such goods in the preceding period, or a pessimistic person may buy a car because his old car broke down.

The simple and crude measure of durable goods purchases used in our test, described in detail in (16), is the number of transactions by each individual family, first, during the year between two succes-

sive surveys (June 1954 to June 1955) and, second, in the six-month period following the second survey (in the second half of 1955). Purchases of one or several automobiles, of large household goods, of a few selected luxury or hobby items, and additions or large repairs to homes were included in the transactions. A scale was constructed ranging from 0 to 5. In the year 1954-55, the average family made approximately 1.5 transactions; in the second half of 1955, about 1.0 transactions.

Calculating the average number of purchases or transactions for each of our groups of families yields data of some interest but does not suffice. Income is an important determinant of durable goods purchases and there are, as shown in the tables in the preceding section, substantial income differences between some of our groups. Therefore, for each attitude group the "expected number of purchases" was calculated, based on the incomes of the members of the group (16). The relation between the actual number and the expected number of purchases provides us with a measure reflecting the impact of attitudes beyond what could be accounted for by income.

In comparing the 1954-55 purchases of pairs of groups which had the same attitudes in June 1954 and different attitudes in June 1955 (for example, the purchases of Group MM with those of MO), should we expect to find any differences? We may answer this question in the affirmative by assuming that the change in attitudes was evenly distributed over the 12-month period, or even occurred overwhelmingly early in the period. Following the second measurement, in the second half of 1955, we should expect Group MO to make more purchases than Group MM since the initial attitudes of

the two groups were different; yet in doing so we disregard the influence of changes in attitudes which may have occurred between June and December 1955.

We find in Table 10A that the change in personal financial expectations affected purchases in both periods. Group OO purchased more than OM, and Group MO more than MM. The differences in the second half of 1955 indicate that attitudes do matter. The differences in the year 1954-55 may be related, in addition, to findings presented in Table 5: Indications for the forthcoming attitude changes were available already in June 1954 (see Approach I), so that we have reason to believe that some attitude changes had taken place shortly after the first measurement.

Regarding changes in the general business expectations (Table 10B), the purchases of the four groups are found to be substantially the same in 1954-55. In the second half of 1955 we find sizable differences in the expected direction. To explain the first finding, we may assume that the information about the changes in business conditions came late during the 12-month period. Evidence supporting this assumption comes from surveys toward the end of 1954, as well as from an analysis of business cycle developments. We may also recall that in this case we found no initial differences between Groups MO and MM (Approach I). This finding also points toward a relatively late shift in attitudes.

According to Table 10 the 1954-55 purchase rates of the initial O groups (people who were optimistic in June 1954) were not higher than those of the initial M groups. But, on the basis of the more comprehensive index of consumer attitudes, those who were O in June 1954

TABLE 10
NUMBER OF DURABLE GOODS PURCHASES

	1954-55		Second Half of 1955	
	Actual	Actual in % of Ex-pected	Actual	Actual in % of Ex-pected
A. Personal Financial Expectations (see Table 5)				
Group OO	1.82	110%	1.20	114%
Group OM	1.45	94%	.93	94%
Group MO	1.69	109%	1.05	105%
Group MM	1.46	93%	.88	92%
B. General Business Expectations (see Table 6)				
Group OO	1.53	99%	1.09	105%
Group OM	1.49	104%	.80	85%
Group MO	1.42	103%	.99	106%
Group MM	1.45	105%	.85	90%

are found to have had a purchase rate of 107% in 1954-55; those who were M, a purchase rate of 102%; and those who were P, a purchase rate of 86% (16, Table 2).

IMPLICATIONS AND CONCLUSIONS

What remains to be done is to discuss the consequences of the findings reported in this paper and to connect them with earlier research in economic psychology. The findings are relevant for the evaluation of survey research, the study of business cycles, and the theory of social behavior.

Survey Research

Survey data have been used frequently for purposes of prediction. When, for instance, surveys conducted with representative samples of the population revealed that intentions to buy durable goods had become more numerous and expectations regarding personal finances and economic developments more optimistic than, say, 6 or 12 months earlier, it was concluded that consumers willingness to buy had become more pro-

nounced. Similarly, in case of deterioration of attitudes, it was concluded that willingness to buy had declined. Such conclusions, based on observed relations between attitudes and subsequent action, may be subject to qualification on two grounds. It has been argued that it is necessary for the attitudes established in a survey to remain unchanged for quite a while—how else could one make valid predictions for the future? Secondly, and this is closely related to the first argument, changes in attitudes have been called irrelevant on the ground that they may have been caused by unreliable measurements or vacillations by respondents. In assessing the potential usefulness of survey research, we have reason to doubt the relevance of both arguments.

We must agree that sometimes the future may appear most uncertain, and then expectations may be based on guesses and people may vacillate. In such a situation, however, so it appears on the basis of the present studies, two consecutive distributions of attitudes are likely to remain similar, and the "prediction" derived from the attitudinal data will consist of the statement that no radical changes in either direction are probable. When, on the other hand, two consecutive distributions show substantial changes, the preponderant direction of change by individuals tends to coincide with the marginal change and the conclusion that consumer sentiment has improved or has deteriorated is warranted. Is it probable that such a conclusion arrived at, say, in June will have some relevance also in July or September of the same year? We have reason to answer this question in the affirmative, provided no unexpected important developments occur in the months immediately following the survey.

Analyzing stability or change of consumer attitudes, we rely on changes in the answers to several questions. By constructing an index of economic attitudes from expectations expressed about personal finances and national business and market conditions, we reduce the impact of occasional unreliable measurements and vacillations. In addition to inquiring about expectations, we also collect data about people's reasons for them. By asking to what people attribute their optimistic or pessimistic opinions—the usual question is “Why do you think so?”—we of course do not receive information about deep-seated motivational factors. Yet information even about superficial reasons helps us in determining at a subsequent date whether later developments are expected or unexpected. If, for instance, a widely expected upturn in business is practically never attributed to price movements and prices change substantially in the months following the survey, we may hold that a new situation has arisen because of which the predictive value of previously expressed expectations has become doubtful. Similarly, an increase in the frequency of intentions to buy automobiles, associated with an expectation of stable auto prices, must be considered with caution if shortly after the survey auto prices are raised substantially.

Experience over the past ten years indicates, however, that major unexpected developments do not occur very frequently in the economic field. Whether they occur, we usually can find out from survey data. When they do not occur, attitudes do have predictive value. Nevertheless, data on attitude changes do not yield a definite forecast about people's future behavior. By supplementing the traditional data about people's ability to buy with data on people's willingness to

buy, the chances of making correct predictions are enhanced but are not made perfect. Unforeseen developments may occur, and people's probable reactions to them cannot be determined in advance.

Business Cycle Research

The literature on business cycles is replete with references to waves of confidence or waves of distrust. Similarly, in analytical studies of specific cyclical developments—whether it is the great crash of 1929 or the upswing of the postwar years—psychological forces which have taken hold of the people have been pointed out frequently. Most commonly, economists have not considered optimism or pessimism by businessmen or consumers as the major causal forces, but have treated them as reinforcing factors compatible with any economic theory of business cycles (3, Chap. 6). Nevertheless, changes in the attitudes of businessmen or consumers have been commonly relied upon to explain how relatively minor causes have had major effects as well as to account for the timing of turning points in cyclical fluctuations. In all these respects, economic literature is usually restricted to a priori considerations or anecdotal evidence.

Past research in economic psychology has made a contribution to business cycle studies. It has demonstrated that changes in the economic attitudes of consumers and businessmen are measurable. Through asking a variety of carefully formulated questions in surveys conducted with representative samples of the population, it has been found possible to obtain information either about the absence of changes in sentiment or about the direction of changes that have taken place (9). Sometimes this information has become available at a time when

it was not particularly useful for the purposes of policy makers in government or business. It confirmed what was generally known or what was widely expected in the light of data on national income, retail sales, production rates, and the like. At other times, however, as will be shown later, observed changes in consumer sentiment have contradicted trends derived from economic data which reflect past activities of the economy.

The present studies go beyond previous contributions in two ways. First, we now add a second measure to our arsenal of tools. In addition to relying on changes in the distribution of attitudes—the presence or absence of substantial marginal changes—we have developed the measure *T* which indicates whether the marginal change coincides with frequent offsetting shifts or results from similar influences which have affected the majority of people. The finding that there tends to be an association between large marginal changes and small *T* rates strengthens our reliance on the first measure.

Secondly, our studies on the origin of substantial changes in consumer sentiment indicate that uniform acquisition of experience is possible and may be effective. It happens that the same information is apprehended by very many people in the same way and leads to similar changes in people's attitudes and actions. Keynes argued that consumer expectations—their income expectations to be precise—are likely to average out for the country as a whole (11, p. 95). Should this be generally true, then economists would rightfully neglect the consideration of consumer expectations. For the economist is not interested in explaining the antecedents of action by individuals. Whether Mr. Smith steps up his rate of consumption because of his optimistic

views or because his income has gone up is not relevant for the economist unless there are many more people who increase than people who reduce their rate of spending. Only in that case will the economy be affected and will the factors bringing forth changes in the behavior of the masses become a concern to economists.

This author has argued before that changes in consumer sentiment usually will not cancel out but will rather resemble a contagious disease which tends to spread to very many people (4). This argument, derived from general sociopsychological principles, was not based on specific studies of the origin of economic attitudes. Now empirical research confirms the occurrence of nonoffsetting changes in economic attitudes. The current studies aim at the specification of circumstances under which it is probable that changes in the attitudes of some people in one direction will cancel the changes in the attitudes of other people in the opposite direction, as well as the circumstances under which this offsetting effect is not probable. The latter is expected to be the case when—and we summarize all too briefly—uniform acquisition of information has taken place.

This answer leads to new questions, the consideration of which will help to clarify the answer itself. First we have to ask: Why do we bother with attitudes; why not rely on the information which resulted in changes in attitudes? If attitude change is the result of economic or political developments about which information is transmitted, we may do better by considering the underlying factors about which factual information is or should be available. This notion must be contradicted. Even complete knowledge of all events and developments

would represent nothing more than a listing of possible stimuli. Even if such a listing is obtained, for instance, through content analysis of news transmitted, we would not know which items of news are salient and how they are apprehended. This can be done only by starting with the attitudes. Knowing of the prevailing attitudes and their recent changes, we may connect them with events and developments because then we are in a position to select the relevant stimuli.

To be sure, there may be exceptions. Events may occur from which we may deduce how they will influence people's attitudes and actions. Developments of overwhelming significance, especially of a catastrophic nature, may belong in this category—although consumer and business reactions to the outbreak of war have not always been correctly assessed. But, neglecting developments which we know *a priori* are salient, it is necessary to analyze further the difficulties encountered in any attempt at deducing the probable attitude changes from observed events.

As our first example we shall consider a situation in which very many families experience income gains. Will this development necessarily result in optimistic attitudes and a high rate of spending? There is evidence that even the effect of income increases on consumer feelings of being better off cannot be predicted with any degree of assurance. Often, though, increases in real income and sometimes even increases of money income did have this effect, but at other times this was not the case. The latter situation occurred, for instance, in 1957. In order to understand it we may recall that in 1954, as discussed before, consumer attitudes improved greatly. At that time, however, income increases were relatively infrequent: In October 1954, 28% of a repre-

sentative urban sample of American families reported income increases; and 27% income decreases. By 1956, the proportion with income increases rose to 32.5%; and by March 1957, to 43.5%. The proportion with income declines was as low as 15.0% both in 1956 and 1957 (the remaining families reported stable income). From 1956 to 1957, however, in spite of the favorable trend in stimuli, the feeling of being better off declined in frequency and consumer optimism also declined. That prices went up in the year prior to March 1957 appears to be only one of the explanations for this observed divergence. Many other considerations—such as the notion that income increases reflected past business improvement which may not continue—must be taken into account and may help to explain the response to income increases, provided the attitudinal data are known. At this point it suffices to say that in March 1957 no valid conclusions could be drawn about the trend of consumer attitudes from observing that income gains greatly increased in frequency and were three times as frequent as income declines.

We may cite briefly a second, somewhat different, example. Suppose that at a given time, discounts offered to buyers become widespread: content analysis of news reveals the establishment of discount houses and the availability of high trade-in values for important consumer goods (as was the case in 1953-54). Perhaps we may then predict that this news will have a favorable effect on consumer demand. But can we also say, without studying consumer opinions and attitudes, that very many people will know of the discounts? Since not all information transmitted an impact on masses of people, in this respect again attitudinal studies remain indispensable.

The most complex example known to us consists of the impact of price trends. Upward price movements are known to generate sometimes the expectation of further price increases; at other times, however, the expectation of price declines. Expected price increases may induce large-scale buying in advance and in excess of immediate needs, but under other circumstances they may be considered a calamity or may restrict discretionary purchases because they are seen to detract from well-being and make it harder to meet necessary expenses. It is not possible to rely on price indices, and it remains necessary to determine what attitudes and expectations result from price changes.

From the conclusion that measurement of attitude changes is indispensable, it does not follow that such measurement always yields significant new insights for business cycle studies. As said before, autonomous attitude changes occur only at certain times, while at other times attitude changes reflect past trends. The argument just presented implies that it is not possible to find out a priori which of the two possibilities prevails. Only after determining that prevailing attitudes do not indicate any significant new developments can one conclude that the attitude measurement did not contribute any new knowledge and that it is permissible to rely on extrapolation of past trends.

Prior to World War II business and government were usually viewed as the sectors of the economy in which changes of economic activity are bound to originate. If business investment or government spending generate higher national income and thereupon consumers add to their expenditures, we cannot assign autonomy to the consumer sector. Three times, however, during the last few years,

autonomously caused changes in consumption were observed. In 1949 a moderate decline in economic activity originating outside of the consumer sector failed to influence the consumers; they maintained their optimistic attitudes and, by increasing their rate of spending, pulled the economy out of the slight recession (4, Chap. 13). In 1951, at a time of rapidly increasing incomes, consumer resentment against price increases and the uncertainty or anxiety with which consumers viewed the cold war resulted in restraint in buying (8). Finally, in 1954 a plateau in economic activity was terminated because of optimistic feelings of consumers who were impressed by the small damage done by the widely advertised recession of 1953, as well as by price stability and the availability of "good buys" (9). Consumer optimism led to further upgrading of consumer possessions for the sake of which broad groups of consumers were willing to allow their instalment debt burden to grow. On the other hand, to mention just one example of rather different developments, late in 1955 consumer sentiment, as well as behavior, reflected improved business conditions and growing income and did not provide new incentives to the economy.¹⁹

From these experiences the conclusion may be drawn that the consumer sector, just as the business sector, may act autonomously as well as nonautonomously.

¹⁹ Consumer attitude surveys, conducted by the Survey Research Center toward the end of 1955 as well as the end of 1956, indicated optimistic sentiment and a fair rate of buying intentions, yet no gains in either respect as against previous measurements. From these findings the conclusion was drawn that "The consumer sector cannot be counted upon to provide any new strength or impetus to the economy." These indications were fulfilled by subsequent economic developments. According to a survey conducted in June 1957, consumer optimism was weakening at that time.

From the point of view of business cycle theory this conclusion does not suffice. The finding that consumers may influence economic activity in an autonomous manner, because new information may change the economic attitudes of masses of consumers in a uniform way, gives rise to further questions.

Two new facts confront us today compared to the situation which has prevailed at earlier times. One is the increased availability of information, and the other the increase in the number of decision makers whose action may influence economic developments. It may be argued that not so long ago news spread slowly and most people were ignorant of happenings outside their limited sphere; today, however, with rapid mass communications, every localized difficulty or incipient tendency becomes widely known. Is this situation conducive to exaggerated and excessive reactions and, therefore, to a spread and snowballing of either inflationary or deflationary tendencies? Regarding the growing influence of the consumer sector, we have to ask: Does it make a difference that several million household units may influence economic activity rather than a few thousand large business units? Both of our questions may be summarized by asking whether too much information received by too many people is detrimental to economic stability.

It has been frequently asserted that expectations may spread rapidly and may become self-reinforcing and self-justifying. One may recall the bank runs in the early thirties. It was then said that banks failed because they were expected to fail. Rumors that a bank was not sound led people to withdraw their deposits; even a strong bank could not resist the cumulative process in which

more and more people participated and made the rumors come true by their action. Or one may think of the vicious circle of the same period: a slight reduction in demand led to cutbacks in production, then to fear of unemployment and the expectation of lower incomes, which in turn made more and more people reduce their expenditures, and incomes declined because of curtailed demand. Contrariwise, inflationary fever was experienced in Germany, France, China, and many other countries. People expected prices to go up, therefore spent their money rapidly, which made prices go up; when people found their expectations fulfilled, they expected further price increases and hastened to hoard goods—that is how the cumulative process of self-realization of expectations was set in motion.

Are these typical instances of human reactions? One may wonder whether the bank runs and the great inflations have been correctly described in the preceding lines. Do we mean to assert that the processes originated in unjustified rumors about the difficulties of banks or about future price increases, and ultimately the unfounded expectations made themselves come true? Even assuming that this may occasionally have happened, is it the common or usual situation? We also have experience with recessions and inflationary movements which were arrested early. Incipient movements of prices or sales are often reversed after a short flurry. During the past ten years in the United States, small recessions and small inflationary movements were the rule rather than the exception. The theory of self-justifying expectations applies to but one type of experience, a rather rare one for that matter, and does not help to explain reversals of expectations.

Although the evidence available at present is fragmentary and not conclusive, it is possible to spell out a set of assumptions which are derived from the notion that the economic thinking of the masses is fundamentally conservative and sane. Cumulative and self-justifying expectations may then be viewed as a form of catastrophic behavior. People resist speculative fever as well as despondency unless their sanity is crushed by a series of repeated shocks. News and rumors which have no solid foundation may be accepted for a short while by some people, but will not sustain action by very many people over prolonged periods.

The basis of mass sanity may be found in the desire to understand the reasons for developments that take place. Unless news is altogether without foundation, some information about what has caused the new developments is transmitted and will be understood. Expectations may originate in two ways: either through projecting past trends or through understanding that there are good reasons for new trends (4, Chap. 4). The first instance, when past price increases generate the expectation of further price increases, is not the only one. Expectations of price increases have been observed to originate at times of stable prices because of the understanding of certain underlying factors which would give rise to a new development. When, then, the expectations are fulfilled—prices have gone up—people may note that what was called for has happened and that the forces are exhausted. Similarly, when expectations are based upon simple projections, people search for an understanding of underlying reasons and, failing to find such reasons, the expectations will not be sustained. Unfounded rumors may influence some people's behavior for

a short time, but in order that expectations of very many people be maintained over long periods, people must understand the reasons for their expectations.

Obviously, the term "understanding" used in the preceding discussion does not involve sophisticated knowledge about economic relationships. It implies a feeling about what leads to what, which is not only within the capacity of broad middle income groups, but is also the prerequisite for their discussing economic events. Such discussion, we stated before, was found to be an essential condition for social facilitation of opinions and attitudes.

Interviews reveal that the large majority of people promptly answer such questions as "Why do you say so?" or "Why do you think so?" following the expression of opinions about the prospective trend of prices or economic activity. To be sure, the reasons are not given in terms of complex theories. Frequently the answers consist of such statements as "Everybody is buying," "There is more demand than supply," "The buyers are in control," "There's lots of money around." Whether or not economic analysts are satisfied with these explanations is unimportant. The people who say so are satisfied; they feel they know why prices will go up. They see a connection between their experience and their expectation.

The understanding of economic developments by broad groups of people seems to involve some stereotyped notions which are fundamentally conservative. "What goes up must come down," is frequently said, or "Trees don't grow to heaven." There are limits to most trends in the opinion of most people. The forces that generate the trends are not thought to be inexhaustible. Often,

the longer a trend continues, the more people are inclined to watch out for reversals.

It is, therefore, possible that in the economic area mass reactions are characterized more by self-regulatory features than by a tendency toward excesses. We may illustrate this notion by describing one of the major developments of post World War II American economy. We refer to the process of "upgrading" consumer possessions and standard of living. In these years gratification of needs—be it for a new car or a larger television set—has commonly resulted in the emergence of further needs, rather than in saturation. The new needs and desires for better products arose among people who were quite well provided for and not among the "have-nots."

These observations were made at a time when the majority of people were confident that their personal financial situation would improve. People had experienced frequent raises in wages and salaries, were satisfied with their occupational progress, and became convinced that they were entitled to expect continuous improvement in their standard of living. Are these expectations and desires without limit? Do they necessarily result in over-buying and subsequent collapse? Consumer purchases are dependent both on felt needs or sentiment and on ability to buy. For many years economic forecasts about saturation with consumer durable goods proved to be wrong because younger, middle, and upper income consumers were willing to enlarge their instalment debt burden for the sake of upgrading their possessions. When in 1955 the proportion of incomes mortgaged for debt repayment rose substantially, government experts called for a regulation of instalment

credit. No regulation was enacted, however, and the experience of the following two years indicated that such a regulation was not necessary. The consumers themselves regulated their financial position. No doubt, some individual families overreached themselves. Many more figured what they considered to be the permissible debt burden on the basis of the income they expected to have in the future rather than on the basis of their current income. (Therefore they and the economy would have been in a difficult situation if the expected income increases had failed to materialize.) But as the slower rate of automobile purchases in 1956 and 1957 has shown, consumers on the whole knew that there were limits to how far they could go in adding to their monthly contractual charges. And they kept in mind that unexpected developments might occur and would have to be provided for through accumulating liquid reserves. The desire to save remained strong at the same time when great emphasis was placed on the improvement of standard of living.²⁰

Thus it is possible that many millions of relevant decision makers, obtaining prompt and similar information about economic developments, add to the stability rather than the instability of the economy. Possibly, the more decision makers there are, the less probable are excessive reactions. It may be argued that dynamic forces may more easily sway an entire group when the group is small—as, for instance, big business—and has the power to control economic activity. With the much more diffuse masses of consumers, early reactions and reversals are more probable.

²⁰ In (5) it has been shown that the importance attached to saving and the desire to save have not declined at all during the ten prosperous years following World War II.

We found that extensive groups of consumers may acquire the same new economic information at the same time and may accordingly change their attitudes and behavior. It is conceivable that occasionally this ability will lead to the vicious circle of deflation or the mass hysteria of runaway inflation. Yet, knowing what it takes to influence the attitudes of masses of consumers, we are in a better position to understand the developments of the past ten years. In this period, consumers contributed to economic stability either by serving as a brake to inflationary trends or by stepping up their demand when the economy needed new incentives so as to emerge from a plateau.

Theory of Behavior

We may reformulate our conclusions without reference to economic developments. While in the preceding discussion we stressed the implications for business cycle analysis, we now turn to behavioral theory.

The finding that acquisition of information may result in change of attitudes means, of course, that cognition may influence affective preferences. There is nothing new in this statement, even though the process has been rarely demonstrated, and the reverse process (affects influencing cognition) has been more frequently emphasized in recent psychological research. The impact of cognition, both on affect and on action, has been shown in a social situation. There is every reason to believe that the similarity of the information acquired by members of the same group, or by people with similar interests, is an essential feature of the learning process here studied and that the similarity of affective reactions is greatly promoted by social interaction. The introduction of the

term "social learning" is justified because social action represents the end result of the process here described. In studying individual behavior we do find people who acquire information about an improvement of economic conditions, thereupon change their attitudes in the optimistic direction, and nevertheless do not make any purchases of durable goods. Conversely, some individuals make such purchases without acquiring new information and without changing their attitudes. What is claimed here relates only to the average behavior of groups of people in the two situations. In other words, we are concerned with social learning which has an impact on the society (or economy).

What can be said about the process of social learning? Is it basically a process of imitation, of following the leader, of keeping up with the Joneses? We find that similar stimuli elicit similar responses. But a statement about similarity of stimuli does not suffice. The information on economic developments reaching the people has innumerable aspects, of which only some become salient and relevant. Our perceptions are selective. Why is it that at certain times the same aspects of information become influential? Here we note the effect of groups on the individual. People belonging to the same face-to-face group or having similar interests (e.g., the occupational groups) have the same kind of apperceptive mass. They are predisposed to select the same information and to understand the information they receive in a similar manner. There also exists pressure toward uniformity both in selecting the information and in understanding it. Finally, there is mutual reinforcement of information among group members. The aspects of information people tend to

discuss in their group are those with which they most strongly agree or disagree, and these conversations tend to result in uniform attitudes. No doubt many individual differences remain even among members of coherent groups. Yet it is the direction of the attitude change which matters. This direction tends to be the same when a "current" ensues which becomes hard to resist.

SUMMARY

We may summarize our model on the origin of changes in expectations in the following Points 1 and 2 and present derivations from the model, made probable by the empirical findings, in Points 3 and 4.

1. Change in expectations is due either (a) to the acquisition of widely transmitted information of a general nature, (b) to personal experiences, or (c) to errors of measurement.

2. At any given time, there will be individuals with whom Variables *b* and *c* operate in one direction, and individuals with whom they operate in the opposite direction. On the other hand, most commonly, Variable *a* is either noninfluential or operates in the same direction with very many people.

3. If Variable *a* is not influential, ag-

gregate changes in expectations will be small because most changes in individual expectations cancel out. Observing substantially unchanged distributions, it is not permissible to assume that most individuals have maintained their previous expectations; on the contrary, it is probable that among individuals there have been frequent changes in both directions. (This statement does not hold good for relatively deep-seated and enduring attitudes, when very small changes in aggregate distributions were observed together with infrequent cross shifts by individuals.)

4. If Variable *a* is effective and substantial, then aggregate changes in expectations will likewise be substantial. This means that under these circumstances contrary effects of variables of Type *b* are suppressed or lessened. Substantial aggregate changes in expectations should then be attributed to variables of the Type *a*. This final derivation from the model, implying causal relations, is suggested (rather than demonstrated) by data on information acquired by different groups as well as by an association between high marginal change rates and low rates of unnecessary turnover.

REFERENCES

1. FERBER, R. On the stability of consumer expectations. *Rev. Econ. Statist.*, 1955, 37, 256-266.
2. GLOCK, C. Y. Some applications of the panel method to the study of change. In P. F. Lazarsfeld & M. Rosenberg (Eds.), *The language of social research*. Glencoe, Ill.: Free Press, 1955. Pp. 242-259.
3. HARBERLER, GOTTFRIED. *Prosperity and depression*. Geneva, Switzerland: League of Nations, 1937.
4. KATONA, G. *Psychological analysis of economic behavior*. New York: McGraw-Hill, 1951.
5. KATONA, G. Attitudes toward saving and borrowing. Part II, Vol. 1. *Consumer installment credit*. Washington, D.C.: National Bureau of Economic Research and Federal Reserve Board, 1957. Pp. 450-487.
6. KATONA, G. Federal Reserve Board committee reports on consumer expectations and savings statistics. *Rev. Econ. Statist.*, 1957, 39, 40-46.
7. KATONA, G. Business expectations in the framework of psychological economics. In M. J. Bowman (Ed.), *Expectations, uncertainty, and business behavior*. New York: Social Science Research Council, 1958. Pp. 59-74.

8. KATONA, G., & MUELLER, EVA. *Consumer attitudes and demand*. Ann Arbor, Mich.: Survey Research Center, 1953.
9. KATONA, G., & MUELLER, EVA. *Consumer expectations*. Ann Arbor, Mich.: Survey Research Center, 1956.
10. KENDALL, PATRICIA. *Conflict and mood*. Glencoe, Ill.: Free Press, 1954.
11. KEYNES, J. M. *General theory of employment, interest, and money*. London: MacMillan, 1935.
12. LAZARSFELD, P. F. The use of panels in social research. In B. Berelson & M. Janowitz (Eds.), *Reader in public opinion and communication*. Glencoe, Ill.: Free Press, 1953.
13. LAZARSFELD, P. F., BERELSON, B., & GAUDET, H. Introduction to *The peoples' choice*, (2nd ed.), New York: 1948. Reprinted in *The language of social research*. Glencoe, Ill.: Free Press, 1955.
14. LIPSET, S. M., LAZARSFELD, P. F., BARTON, A. H., & LINZ, J. The psychology of voting: An analysis of political behavior. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. II. Cambridge: Addison-Wesley, 1954. Pp. 1124-1177.
15. MACCOBY, ELEANOR E. Pitfalls in the analysis of panel data. *Amer. J. Sociol.*, 1956, 61, 359-363.
16. MUELLER, EVA. Effects of consumer attitudes on purchases. *Amer. Econ. Rev.*, 1957, 47, 946-966.

(Accepted for publication July 8, 1957)

APPENDIX

The technique used for the purpose of measuring the level of economic information of the panel consisted of handing a printed card to each respondent with the following instruction: This card contains some answers we received when we asked people about what had happened during the last twelve months in the American economy. Please check those items which you agree have happened.

There were three columns on the card, one of which had to be checked for each statement. The columns were labeled "Happened," "Did Not Happen," and "Don't Know." Some of the statements were true, others were false. The answers were scored for correctness; people who checked "Don't Know" were taken as not having correct information.

The seven statements given to the respondents in June 1955 read as follows: What happened during last 12 months?

1. Business conditions improved.
2. The cost of living was stable.
3. Unemployment increased.
4. People have less money to spend than a year ago.
5. Stock prices went up.
6. Defense spending by the Government declined.
7. The Federal budget was balanced last year.

Altogether, 27% of the sample answered 0, 1, or 2 items correctly ("Hardly any correct information"); 53%, 3, 4, or 5 items ("Some correct information"); and 20%, 6 or 7 items ("Much correct information"). The mean number of correctly answered items was 3.73. Since the economy improved greatly in the year prior to June 1955, correct information refers to favorable developments.

The percentage of correct information received was highest regarding the cost of living (Item 2) and was relatively high also regarding four other items. Least was known about stock prices (Item 5) where the majority checked "Don't Know,"

and about defense spending (Item 6) where the frequency of "Happened" and "Did Not Happen" answers was quite similar.

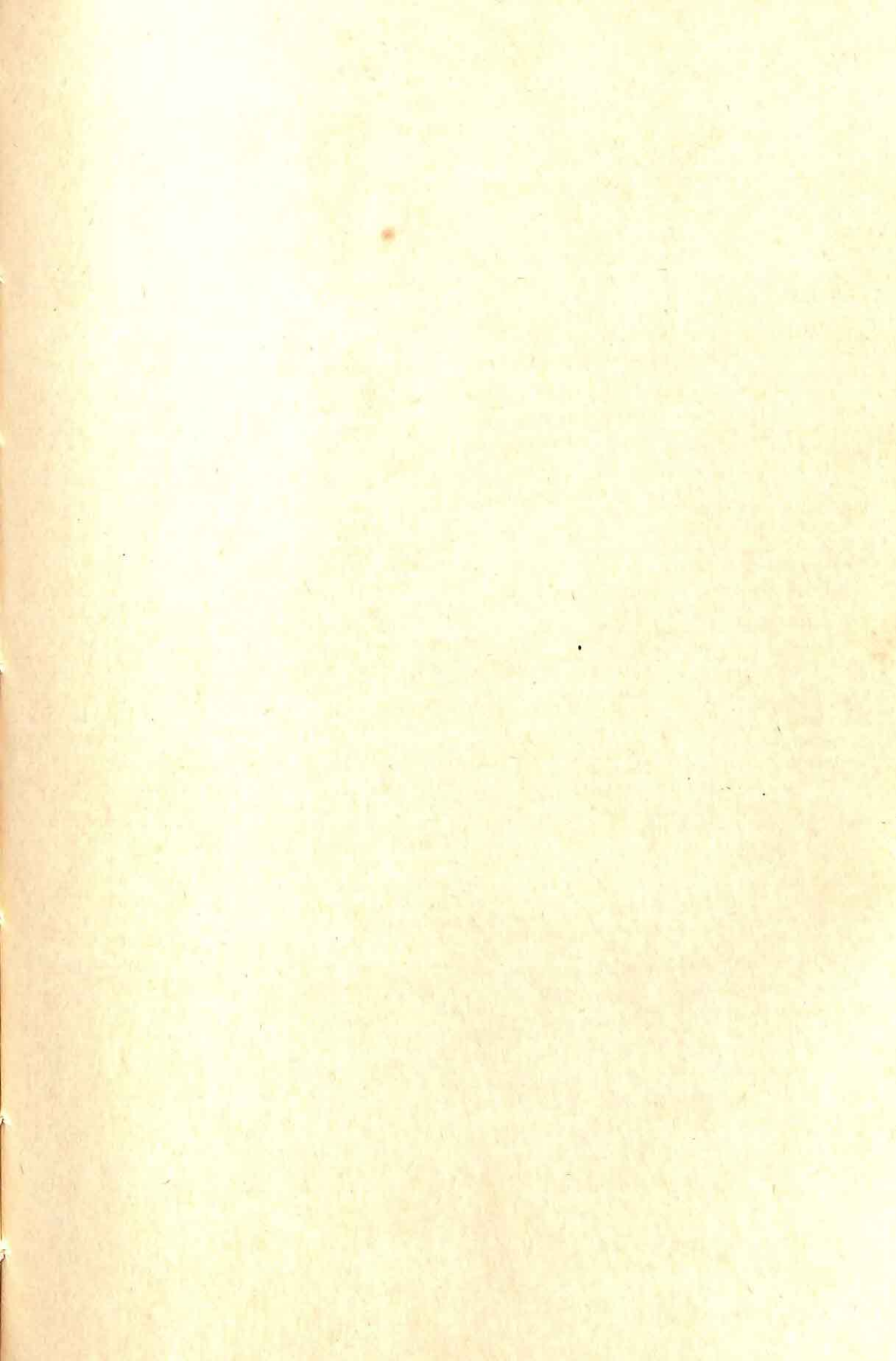
Questions on economic information, identical in form but different in content from the previous ones, were also asked in March 1957. The seven statements, about which respondents were asked to say whether they had or had not happened during the past twelve months, were the following:

1. Industrial production declined.
2. The cost of living rose somewhat.
3. Unemployment increased.
4. Stock prices went up sharply.
5. On the average incomes increased.
6. Interest rates rose.
7. Fewer cars were sold in 1956 than in 1955.

The correct answers were that Items 1, 3, and 4 had not happened and that Items 2, 5, 6, and 7 had happened. As can be readily seen, some correct answers represent favorable developments and some correct answers unfavorable developments in the economy.

Altogether, 22% of the sample answered 0, 1, or 2 items correctly (hardly any correct information); 46%, 3 or 4 items; and 32%, 5, 6, or 7 items (Much correct information). The mean number of correctly answered items was 3.74.

Detailed analysis shows that the great majority of people answered correctly the questions about the cost of living, about incomes, and about interest rates (Items 2, 5, and 6). Regarding each of the other four items, fewer than half of the respondents gave correct answers. (The others either said that they didn't know or gave incorrect answers.) As mentioned above, the frequency of incorrect information regarding two items was particularly notable. The majority of lower and middle income people thought (incorrectly) that unemployment had increased; in all income groups less than half of the people knew that fewer cars were sold in 1956 than in 1955.



Psychological Monographs: General and Applied

Factor Analytic Study of Visual Memory¹

RAYMOND E. CHRISTAL

Wright Air Development Center

MEMORY plays an important role in one's success in school, on the job, and in dealing with one's fellow beings. It is not surprising that psychologists have devoted a considerable proportion of their total effort to the study of this process. It is somewhat discouraging, however, to realize how little we know about it.

There is a growing body of evidence that memory is not a single unitary function. Evidently this has long been suspected, for before any attempt was made to separate and classify memory abilities, experimental psychologists were referring to incidental memory as if it were different from direct memory, immediate recall as if it were different from delayed recall, memory for meaningful material as if it were different from memory for nonsense material, memory for form, memory for color, auditory memory, memory span, and memory for position in space.

The present study was designed to investigate the factorial structure of visual memory. More precisely, it was designed to identify and define factors corresponding to the ability or abilities of individuals

to recall stimuli presented in the visual modality. There have been a number of studies in which auditory memory and memory span factors were identified but these are not relevant to the present investigation which is restricted to tests administered in the visual modality. In this respect there is only one well-established memory factor. This is the factor M that was extracted in Thurstone's analysis in 1938 (12). It has been called rote memory, memory, associative memory, and paired associates memory; but in all cases it seems to be the same. It is best defined by tests requiring the formation of associations and will be referred to in this study as the associative memory factor.

One other memory factor in the visual memory area is considered by some psychologists to be established. This factor, called visual memory, was identified in four studies conducted by the Air Force (8, pp. 261-268, 408-418). It seems to represent the ability to recall past visual experience. However, in two of these studies the factor was extremely weak and its interpretation questionable, while in the other two studies the tests identifying the factor were limited to measures of map memory. It was hypothesized that the ability is general in nature and that it would be found in tests other than those involving maps as subject matter (9, p. 228).

In 1937, Carlson (3) extracted what appeared to be three visual memory factors from a battery

¹This report represents a condensation and rewriting of a dissertation submitted to the University of Texas in 1956 as partial fulfillment of the requirements for a doctoral degree. Benjamin Fruchter served as chairman of the committee. It also is based in part on work done under ARDC Project No. 7719, Task 17008, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas.

of 35 tests involving visual and vocal recognition of words. In an Air Force re-analysis of this study (9), intercorrelations were averaged to reduce dependency and only one factor was extracted. As French points out (6, p. 248), the evidence for a visual memory factor is quite thin. There has been no cross identification between these various analyses. In two of the Air Force studies the factor is extremely weak and its nature unclear. In the two other Air Force studies the factor could be a content factor specific to word lists. There seems to be little justification for accepting a visual memory factor until tests from these studies have been included in a single analysis.

Kelly (10), in 1954, completed a study in which he hypothesized a visual memory factor but was unable to demonstrate this factor in his results. He did manage to separate memory for meaningful material and memory for meaningless material into what appears to be two separate factors. These need to be confirmed by additional studies.

HYPOTHESES

The hypotheses of this study take the form of postulated factors. Nine new factors, which can be subsumed under five major headings, are postulated in the memory area. These are listed below.

Visual Imagery. This is postulated to be a factor which will account for memory through use of visual imagery. It is not clear whether a single first-order factor should account for this variance. The visual image can be conveniently broken into three categories: color, form, and position of the elements within an image. Memory for each of these categories can be thought of independently of the remaining two. One can visualize the color and/or shape of a book and yet be unable to recall its position when last seen. A file clerk may be able to recall seeing a bulletin in a particular section of a particular file drawer without being able to recall its color or shape. Tests measuring each of these three types of memory were designed for the present study. If three first-order factors corresponding to these tasks emerge from the study, the in-

tention is to allow the axes to go oblique, determine the intercorrelations among the factors, and perform a second-order solution in an attempt to locate a second-order factor corresponding to memory through use of visual imagery.

Incidental Memory. This factor would account for memory for things and events encountered incidentally as opposed to memory for things and events which individuals remember through conscious effort.

Shellow (11) made a summary of studies in this area in 1923. She offered three criticisms which apply to practically all of the work done so far under the title of incidental memory. In addition to the tests being too complicated, she indicates that in most cases no care was taken to insure the perception of all of the material upon which questions were to be asked later, so that the tests resolved themselves into measures of observation rather than of incidental memory. In other instances, she states, the subjects were informed that they would later be asked to recall the material which was presented to them, and in this way the tests became measures of direct memory. In her own study she administered a group of tests measuring incidental and direct memory for pictures, names, sequences of events, colors, letters, shapes, position in space, and prose materials. She found that the correlation between summation scores for incidental memory and direct memory was .34, which was somewhat higher than the intercorrelations among tests of incidental memory for various materials or among tests of direct memory for different kinds of material.

Thurstone (13, p. 201), in 1910, asked whether incidental memory is the same ability as the ability to memorize intentionally. Evidently, as time passed, he became more and more convinced that the two were not the same, for in 1946 he stated: "... there probably exists also a separate ability for what is called incidental memory, namely, the ability to recall past experience without previous intention to recall it" (15, p. 108).

Delayed Recall. This factor would represent the ability to recall things and events after a lapse of time as opposed to recalling things and events of the more immediate past. Postulation of such a

factor implies differential loss in retention with the passing of time.

Although many investigators have discovered a rapid loss of retention during the first few hours after the learning session, there are very few studies in which an attempt has been made to determine whether some individuals are more efficient than others in retaining material over periods of time. Forlando and Bunschwig (5), in 1938, reported a study in which 156 children were given immediate and delayed recall tests after limited learning of poetry material. Approximately 40 per cent showed a higher score on the delayed than on the immediate recall form. In another study, Anastasi (2) found markedly different patterns of correlations for two memory tests which were very similar except for the factor of delay. She concluded that certain devices which might facilitate the recall of a series of words immediately after their presentation might be of little value when recall is tested after 48 hours. In an Air Force factor analysis (9) performed on scores from aviation cadet subjects, a test involving the recall of orally presented tactical plans after a delay of two hours was found not to be loaded on the paired associates (associative) memory factor. It was hypothesized that this test measured either an auditory or a delayed memory factor or both.

Memory for Position in Temporal Succession. This factor would account for differences in the ability of individuals to remember the relative position of events in time.

Research in psychology indicates that recall of order within a series is, to some degree, a matter of intraserial association. To the extent that this is true, one would expect tests designed to measure memory for position in temporal succession to have their variance accounted for by the associative memory factor. Thurstone (15, p. 108) indicated in 1946 that there may be a factor concerned with memory for temporal sequence which is distinct from the associative memory factor.

Memory Content Factors. These factors would account for special abilities of individuals to remember certain types of content. Memory for color and form were mentioned earlier under Visual

Imagery. Factors also are postulated to account for memory for words and memory for numbers.

Thurstone (15, p. 108) indicated his belief that the memorizing factor transcends the nature of the content and so becomes applicable to numerical, spatial, or verbal material. Humphreys (9, p. 223), on the other hand, states that memory factors can probably be multiplied almost indefinitely by relatively slight changes in the format and content of memory tests. Vernon (16, pp. 152-153), Achilles (1, p. 24), Grant (7, p. 514), and Anastasi (2, p. 51) tend to agree with Humphreys.

DESCRIPTION OF THE TESTS

This section is devoted to a verbal description of the tests included in the battery. A familiarity with the test titles and content is prerequisite to an understanding of the discussion to be presented later.

Experimental Tests

*Color Discrimination.*² This test presents 30 groups of colored drawings. Each group contains three symbols, numbers, or letters, all three of which are colored blue, yellow, red, green, or brown. However, in each group one of the three is colored a different shade than the remaining two. The examinee is requested to identify the "off-shade" symbol, number, or letter in each group of three. The Color Discrimination Test is not a measure of memory, but it provides study material for the Color Recall Test. It was designed to be administered as the first test in the battery and before the examinees discover they are taking a series of memory tests.

15. *Color Recall.* This test was designed to be administered immediately after the Color Discrimination Test. It consists of the same groups of drawings printed in a neutral gray. The examinee's task is to recall the color of the groups as they appeared in the Color Discrimination Test. The Color Recall Test was designed primarily to measure memory for color, incidental memory, and memory for form.

29. *Color Memory.* The examinee is shown a

² The numbers before the test titles are the identification numbers used in Fig. 4, Tables 1 through 5, and in the factor descriptions to be presented. No number is assigned to the Color Discrimination Test since this variable was not included in the final factor solution.

series of 20 large flash cards, each containing a colored drawing of a common object, at the rate of one card each two seconds. He is asked to remember the colors of the objects. In some cases these objects are colored as one would find them in reality. In most cases, however, the objects are colored inappropriately. For example, the set includes pictures of a green horse, a red cow, and a yellow gun. The examinee later is given the names of the objects and is asked to identify their colors.

28. *Position in Succession.* The examinee is requested to recall the order in which the flash cards in the Color Memory Test were presented. In each question, the examinee must identify which of two cards was presented first. This test was designed to measure incidental memory and memory for position in temporal succession.

14. *Colored Form-Word Association.* The examinee is asked to associate a series of 24 colored forms which are paired with simple English words. After the study session he is presented the forms and asked to identify the words with which they were associated. This test was included as a measure of memory for color, associative memory, and memory for forms.

18. *Position Recall I.* The examinee is asked to study four pages, each of which contains 12 drawings. He then is asked to recall the position of the drawings as they appeared on the study pages. This test was designed to measure memory for position in space.

19. *Position Recall II.* In this test the examinees are requested to recall on which of the four study pages each of the 48 drawings in the Position Recall I Test appeared. It was designed to measure memory for position in temporal succession. Since the examinees were not told that they would have to recall the pages on which the drawings appeared, this test is also a measure of incidental memory.

17. *Picture Detail Recall.* In this test the examinees are asked a series of unexpected true-false questions concerning the content of the drawings in the Position Recall I Test. It was designed to be administered approximately five hours after the Position Recall Test so that it would serve as a measure of delayed recall. It also is a measure of incidental memory and memory for form.

24. *Number-Word Association.* In this test the examinee is asked to associate one- and two-digit Arabic numbers with simple English words. He later is shown the words and asked to identify the correct numbers. This test was designed to measure associative memory, memory for words, and memory for numbers.

23. *Position Memory.* This test was designed to be administered unexpectedly approximately four hours after the Number-Word Association Test. The examinee is requested to recall the

position on the page of the number-word combinations as they occurred in the original test. It was constructed to measure delayed recall, incidental memory, and memory for position in space.

22. *Space Memory.* The examinee is asked to study a page containing 5 squares, each of which contains a symbol in each of 5 sections. He later is requested to identify the symbol that was located in each of the 25 sections. This test was designed to measure memory for position in space, memory for form, and associative memory.

21. *Pattern-Word Association.* The examinee is asked to associate 20 patterns with simple English words. He later is presented with the patterns and asked to identify the correct words. This test was designed to measure associative memory and memory for form.

27. *Pattern Detail Memory.* The examinee is shown a series of 15 patterns formed with straight lines. He later is asked to select each of these patterns from among five similar ones. The test was designed to measure memory for form.

20. *Sequence Memory.* The examinee is asked to recall the order in which the tests in the operational Airman Classification Battery were administered to him approximately three days previously. In each question, the examinee is requested to select the one of a pair which was given first. In order that he may recall the tests by their titles, a sheet is provided which has a sample item from, or description of, each of the tests in the battery. This test was designed to measure delayed recall, memory for position in temporal succession, and incidental memory.

16. *Word Association.* The examinee is asked to associate 30 pairs of English words. Later he is presented with one member of each pair and asked to identify the other member. This test was designed to measure associative memory and memory for words.

25. *Number Association.* The examinee is asked to associate pairs of one- and two-digit Arabic numbers. He later is presented with one number from each pair and asked to identify the number with which it was associated. This test was designed to measure associative memory and memory for numbers.

26. *Turkish-English Association.* The examinee is asked to associate 20 Turkish words with their English counterparts. He later is shown the Turkish words and asked to identify the English words with which they were associated. This test was designed to measure associative memory and memory for words.

Reference Tests

12. *Memory for Landmarks, BI 510A.* The examinee is asked to study diagrams of geographical features (such as rivers and lakes)

which are paired with names. Later he is asked to match the names with the correct diagrams of geographical features.

1. *Word Knowledge, BI 602A.* This test consists of items measuring vocabulary. In each item, the examinee is to select the word or phrase which is most nearly synonymous with the stem word or phrase.

2. *Arithmetic Reasoning, BI 201B.* This test consists of arithmetic problems presented verbally in such a manner that the examinee must determine his own method of solution. The examinee does not need formal mathematics training in order to solve the problems.

3. *Dial and Table Reading, BP 622-621A.* This test consists of three parts, all highly speeded and administered with different time limits. In Part I the examinee has to determine readings on 57 separate dials which are similar to those found on aircraft instrument panels. These dials indicate such phenomena as altitude, air speed, amperage, voltage, temperature, fuel-air ratio, rpm, and so forth, and are of different shapes and sizes. In Parts II and III the examinee has to determine certain information from tables of numbers. For example, in one section he is requested to determine ground speed from tables showing the interrelationship of ground speed, air speed, and wind velocity.

4. *Numerical Operations II, CI 702B.* This is a test consisting of simple subtraction and division problems. The problems and alternatives are printed directly on an IBM answer sheet. Part I of the Numerical Operations Test, which consists of addition and multiplication problems, is administered as a pretest, but is not scored.

5. *Aviation Information, BI 101B.* This test consists of 5-alternative items covering information about the structure and function of airplanes, general aviation information, flying information, and meteorology.

6. *Background for Current Affairs, BI 102B.* This test consists of items concerning information about current events, modern history, geography, and economics. In five of the items the examinee is requested to identify various areas on a map of the world.

7. *Electrical Information, BI 901B.* This test consists of a series of questions about electricity. The examinee is required to demonstrate a knowledge of electrical principles and to read electrical diagrams.

8. *Mechanical Principles, BI 903B.* This test consists of two types of items: mechanical principles and mechanical movements. The items dealing with mechanical principles measure the understanding of such things as leverage, rotation, stress, and transformation of motion. The remaining items require the examinee to comprehend the interrelationship of moving parts

within a machine. All of the items are in pictorial form.

9. *General Mechanics, BI 902B.* This test consists of items covering information about the structure, function, disfunction, and repair of machines. Most of the items concern automobiles, motorcycles, bicycles, and other devices familiar to the average American boy. The items are verbally presented.

10. *Tool Functions, BI 904B.* There are two types of items in this test. In the first type a picture of a tool is given and the examinee must identify the correct use of the tool. In the second type the examinee must select the one of five pictured tools that is used for a certain task or job.

11. *Speed of Identifications, CP 610A.* This is a highly speeded test in which the examinee is asked to observe an aircraft and then identify it from among five similar ones. In each case the planes which serve as alternatives have been rotated in a haphazard manner. The test is in pictorial form and is printed directly on an IBM answer sheet.

13. *Pattern Comprehension, BP 803C.* In this test a flat pattern layout, showing the outline in solid lines and edges along which the folds are to be made in dotted lines, is presented beside an isometric drawing of the three-dimensional object that would be formed by folding the pattern correctly. Certain edges of the folded object are lettered, while edges and fold lines on the flat pattern are numbered. These numbers represent the questions and the examinee's task is to determine the corresponding points on the folded three-dimensional object.

30. *Age-in-years.*

31. *Education-in-years.*

SPECIAL PROBLEMS RELATING TO TEST CONSTRUCTION

The most time consuming part of this study, by far, was the construction, try out, revision, and final printing of the tests. This stage also can be considered to be the most important for two related reasons: (a) the only hope of identifying an hypothesized factor from a matrix of intercorrelations is that two or more of the tests in some degree measure the type of variance hypothesized and (b) if an hypothesized factor is not extracted, no statement can be made as to its nonexistence unless it can be demonstrated convincingly that the tests included in the battery provided ample opportunity for the particular function to be represented. This section is devoted to a discussion of some of the problems encountered in the development of tests for inclusion in the battery. This information should prove helpful to anyone desiring to make further studies in this area. It also will

aid the reader in understanding some of the factor loadings reported later.

Tests Measuring Memory for Color

Only one test could be located, in a search of the literature, which involved memory for color and which could be adapted to group administration. This test, called "Color-Word" (6, p. 101),³ consists of solid-colored forms paired with simple English words. After a form of this test had been constructed for the present study, it was administered to a small group. The individuals in this group were questioned after testing about the methods used in studying and recalling the colored form-word combinations. Results of these interviews indicated that this test probably would be a poor measure of visual recall of color. All of the examinees admitted using verbal associations in "fixing the pairs in their minds," and many of those interviewed stated they had no visual experience in recalling the colored forms—to them it was strictly a matter of verbal associations. For example, the first item in the test presented a blue cross paired with the English word "feel." One of those interviewed stated he associated the "blue cross" with a type of hospitalization insurance. He then associated the words "blue cross" with the word "feel" by saying to himself "you'll be glad you have Blue Cross when you feel sick." Thus, when in the recall portion of the test he was asked to identify the word which was paired with the blue cross, he was able to do so by means of verbal associations. This same individual was asked to name the color of a two-cent stamp, which he did without hesitation. When asked how he recalled this color, he replied "I can see it in my mind's eye. I get a picture of it."

It was obvious that some technique had to be developed to keep examinees from using verbal associations if a visual memory for color factor were going to be defined. One of the techniques used was that of presenting the study materials so rapidly that there was no time for making complex verbal associations. The Color Memory Test, which was designed to utilize this technique, consisted of colored drawings of common objects on large flashcards which were shown to the examinees at a rate of one every two seconds. In the interviews which followed an experimental administration of this test, the examinees indicated that in order to recall the color of the objects presented it was necessary for them to "have a visual experience." Only in one or two instances did any of the examinees state that verbal associations were used, and in

these cases the associations were ready-made and not forced. For example, one of the examinees stated that he remembered the green horse because when he was in high school a mule had been painted green and white (school colors) and taken to a football game.

Another test was worked out which appeared to minimize examinee use of verbal associations to a large extent. This was the Color Recall Test which combined memory for color with incidental memory. It required the examinees to recall the color of symbols in the Color Discrimination Test. Since the examinees did not suspect that they would have to recall these colors, there was no reason for them to make conscious associations. In order to force examinees to take a good look at the groups in the Color Discrimination Test, it was necessary to make the discriminations very difficult and to threaten examinees with a severe penalty for incorrect responses.

Tests Measuring Memory for Form

There was a strong tendency for examinees to use verbal associations to remember form in much the same manner as they utilized this crutch in recalling color. For example, the Pattern-Word Association Test, which was designed

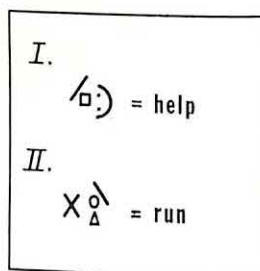


FIG. 1. Two example items from the study section of the Pattern-Word Association Test.

to be a measure of form memory, contained figures composed of circles, lines, curves, dots, and so forth, paired with English words. Two of these pairs which served as stimulus material are illustrated in Fig. 1. Examinees reported a strong tendency toward using verbal associations in recalling pairs such as these. For example, one examinee, in the first item in Fig. 1, visualized the pattern as a man's head whose mouth is "skewed up as if he were yelling 'help'." In the second item he saw a man (the delta and small circle) running from a bear (the X) into a stone wall (the slash).

The Pattern Detail Memory Test, which was designed to be a measure of form memory, is composed of a number of straight line drawings such as those illustrated in Fig. 2. Again, there

³ Called Colored-Form Word Association Test in this paper.

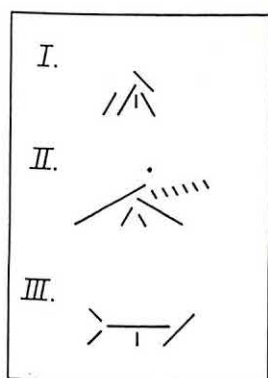


FIG. 2. Three example items from the study section of the Pattern Detail Memory Test.

was the tendency for examinees to use verbal associations in remembering the patterns. For example, one individual saw these drawings as a telescope, a cannon firing six shots, and an arrow flying toward a leaning tree. However, the alternatives to the recall questions in this test were constructed to minimize the value of such associations. For example, Fig. 3 shows choices which served as alternatives to the three stimulus drawings presented in Fig. 2. It can be seen that the first group of alternatives presents

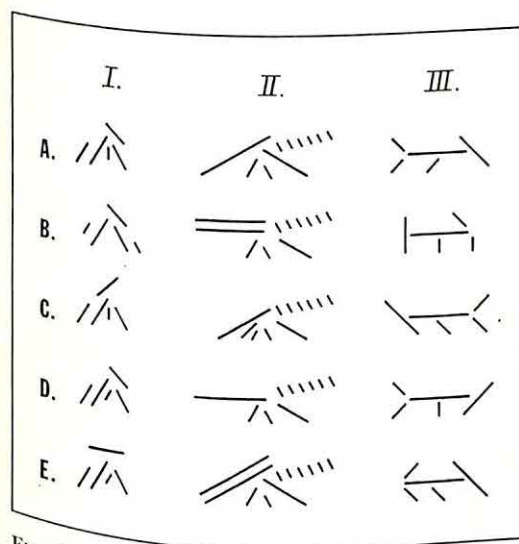


FIG. 3. Alternatives for three items in the Pattern Detail Memory Test. (See Fig. 2 for corresponding stimulus materials.)

five "telescopes" and the second group of alternatives presents five "cannons," each of which is firing "six shots." In the last item the individual examinee mentioned was able to elimi-

nate two of the alternatives. He reported, however, that, for most of the items, he had to depend upon general feeling, guessing, or visualizing.

Tests Measuring Memory for Position in Space

No special precautions were taken to restrict the examinees' use of verbal associations on the memory for position in space tests. It will be seen later that this was unfortunate.

It will be noted that each of the tests in this area measures memory for position in two-dimensional space. However, this was not considered to be a serious problem. A number of spatial and visualization tests which are printed in two dimensions have been demonstrated to have unique validity for jobs requiring manipulation of objects in three-dimensional space. For example, spatial-type tests have been shown to be uniquely valid for predicting success in learning to fly an aircraft (8, pp. 269-296, 477-510).

Tests Measuring Incidental Memory

There were two main problems encountered in the construction of tests to measure incidental memory: to make sure that the examinees did not suspect that they were going to be required to remember the materials used and to insure that the tests had satisfactory reliability.

The solution of the first problem was not as simple as it might appear. Once a group has been "tricked" by one measure of incidental memory, they become extremely suspicious of later tests. Several devices were used to insure that the tests in the present battery would measure this hypothesized ability.

The first device was simply that of assigning one of the measures of incidental memory the first position in order of test administration so that it was given before the examinees discovered they were being subjected to a series of memory tests. It has been pointed out already how the Color Recall Test was placed in this position. The Sequence Memory Test used a similar approach in that it asked the examinees questions about something that happened three days prior to the administration of the memory battery.

Another method utilized to keep examinees from suspecting the nature of the incidental memory tests was to keep them so occupied with one task that they would not have time to study aspects of the test materials used for measuring incidental memory. For example, in the Color Memory Test the examinees were exposed to a series of 20 colored pictures at the rate of one each two seconds. They were told that the pur-

pose of this test was to measure their memory for the colors of the objects shown. After they took the Color Memory Test they were asked unexpectedly to identify the order in which the pictures were presented. In the Position Recall I Test the examinees were permitted to look, for a period of one and one-half minutes, at 12 pictures on each of four printed pages and were told that they would be tested later for their ability to remember the position on the page of each of these 48 pictures. Later the examinees were asked unexpectedly in the Position Recall II Test to identify *on which of the four pages* each of the pictures appeared. Still later, in the Picture Detail Memory Test, they were asked to answer a series of true-false questions about specific content of the pictures.

There were some problems in constructing incidental memory tests having satisfactory reliability. Unfortunately, the devices used to keep the examinees from suspecting the nature of these measures tended to reduce their reliability. For example, one would expect a great deal of chance variance on the Position in Succession Test, which requested the examinees to recall the order in which the colored pictures were presented at the rate of one every two seconds. In this case, the total study time for the test was only 40 seconds. Other devices tended to withdraw the examinees' attention from the subject matter of the incidental memory measures. For example, in the Color Discrimination Test the examinees had no reason to pay attention to the color of the symbols. In general, nothing could be done to improve the reliability of these tests except to increase their length.

Tests Measuring Delayed Recall

The proposed existence of a delayed recall factor rests upon the assumption that there is a differential loss in retention among individuals, e.g., individuals who are equally capable of recalling materials immediately after their presentation will differ in ability to recall these materials after a lapse of time. The big question here is how much delay must there be before such differential loss occurs to a significant degree. There are reasons to believe that differences among individuals may develop in a matter of hours. Ebbinghaus (4) and many other investigators have shown that loss in retentivity occurs very rapidly during the first few hours after exposure to the stimulus materials. In the previously mentioned studies conducted by Anastasi (2) and the Air Force (9), there were indications that a delayed memory factor might be extracted from tests involving only a few hours delay. The present study included three measures of delayed recall. One of these tests involved a delay of approximately three days.

The other two involved delays of approximately four and five hours.

Tests Measuring Memory for Position in Temporal Succession

No special problems were encountered in the development of tests to measure memory for position in temporal succession. The tests which were produced to measure this factor were not highly reliable, but were considered satisfactory for the present study.

Tests Measuring Associative Memory and Memory Content Factors

The associative memory factor has been identified in many previous analyses. The associative memory tests in the present study were selected first on a basis of their loadings in these analyses and second to make possible the extraction of content factors, provided they exist. For example, two tests involving numbers and five tests involving words were included. A new test, the Pattern-Word Memory Test, was constructed so that two of the associative memory measures would involve memory for form.

TEST ADMINISTRATION AND PROCESSING

In this section is a discussion of the experimental and final administrations of the memory battery, a description of the sample, a listing of the test reliabilities and distribution statistics, an outline of the testing schedule, and a brief discussion of the data processing procedures.

Experimental Administrations

Tests in the memory battery were administered, one or two at a time, to samples of 300 to 400 basic airmen. The purposes of these several administrations were: (a) to evaluate the study time allowed for each test, (b) to discover the types of questions asked by examinees, (c) to get estimates of the amount of recall time necessary for 95% of the examinees to finish each test, (d) to give the administrators experience with the new tests, (e) to observe the reaction of examinees to the measures of incidental mem-

ory, (f) to furnish distribution statistics for evaluating difficulty levels of the tests, and (g) to furnish data for computing reliability coefficients. Some of these items will be discussed briefly in the following paragraphs.

Determination of time limits. All of the tests in the memory battery were designed to be "power" measures. That is to say, they were designed to elicit answers under unspeeded conditions. Time limits for the recall portions of the tests were established by the 95% finish criterion, that is, by recording the time it took for 95% of the individuals in the experimental groups to complete each test. The problem of establishing time limits for the study sections of the memory tests was somewhat more complex. If everyone were given all of the time he wanted on the study sections, the tests would tend to become measures of persistence rather than of memory. Practically anyone could make a perfect score on most of the tests if he were given sufficient time and applied himself. In general, the time limits for the study sections were established by observing when the examinees began to get restless and by asking individuals when they felt they had had enough time. A check on the reasonableness of these limits was made by observing the means and standard deviations of the experimental groups' score dis-

tributions. Where the means were too high or too low, a change was made in the study time limits.

Distribution statistics and reliability estimates. Answer sheets from the experimental administrations were machine scored using formulas involving the conventional corrections for guessing. Reliability estimates and test distribution statistics computed from these data are reported in Table 1. The corrected odd-even reliability estimates reported in this table show that even the Position in Succession Test, which was somewhat delayed, incidental, and based upon a 40 second study time, was moderately reliable. The Pattern Detail Memory Test, which was one of the most difficult tests in the battery, had the third lowest reliability. When the study time for this test was increased from five minutes to eight minutes, its estimated reliability, rose from .66 to .71. The Picture Detail Memory Test, which had a relatively low reliability, was composed of 33 true-false questions, incidental, and administered approximately five hours after the examinees were exposed to the stimulus materials.

Final Administration

Order of test administration. The time schedule presented in Table 2 was based upon experience gained during the experimental administrations. Because of the nature of the study it was not possible to set up a flexible schedule

TABLE 1
DISTRIBUTION STATISTICS AND RELIABILITY ESTIMATES FROM THE EXPERIMENTAL
ADMINISTRATIONS OF THE MEMORY BATTERY TESTS

No.	Variable	No. Items	Study Time (Mins.)	N	M	SD	r_{11}
14	Colored Form-Word Association	28	10	370	5.3	5.3	.77
15	Color Recall	30	—	368	4.1	3.7	.68
16	Word Association ^a	—	—	—	—	—	—
17	Picture Detail Recall	33	—	370	25.0	4.8	.63
18	Position Recall I	48	6	364	25.4	10.0	.92
19	Position Recall II	36	—	367	13.5	5.8	.80
20	Sequence Memory ^a	—	—	—	—	—	—
21	Pattern-Word Association	21	8	353	4.5	4.4	.74
22	Space Memory	25	5	353	11.1	6.6	.92
23	Position Memory	20	—	354	8.5	6.9	.83
24	Number-Word Association	20	8	355	11.0	6.2	.85
25	Number Association	15	8	354	6.6	4.7	.81
26	Turkish-English Association	20	8	353	9.9	6.0	.87
27	Pattern Detail Memory ^b	15	5	353	3.1	3.4	.66
28	Pattern Detail Memory ^c	15	8	370	4.8	3.7	.71
29	Position in Succession	40	—	363	12.8	6.7	.58
	Color Memory	20	.67	370	14.2	3.9	.72

^a Being reprinted during time of experimental administrations.

^b First administration.

^c Second administration after revision of study time limits.

TABLE 2
SCHEDULE FOR FINAL ADMINISTRATION OF THE MEMORY BATTERY

No.	Variable	Adminis- trative Directions	Study Time (Mins.)	Recall Time	Overall Time
15	Color Discrimination	5	—	12	17
24	Color Recall	5	—	13	18
	Number-Word Association	4	8	11	23
	Position Recall ^a	7	6	34	47
	(Break)				
14	Colored Form-Word Association	4	10	15	29
27	Pattern Detail Memory	4	8	11	23
26	Turkish-English Association	4	8	13	25
	(Lunch)				
20	Sequence Memory	5	—	15	20
16	Word Association	4	9	14	27
29	Color Memory	5	.67	7	13
28	Position in Succession	4	—	11	15
23	Position Memory	5	—	11	16
	(Break)				
22	Space Memory	6	5	19	30
25	Number Association	4	8	11	23
17	Picture Detail Recall	5	—	10	15
21	Pattern-Word Association	5	8	15	28

^a Provided study materials for Position Recall I and II (Variable No. 18 and 19).

TABLE 3
INTERCORRELATIONS AND DISTRIBUTION STATISTICS OF THE

No.	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Word Knowledge	—												
2	Arithmetic Reasoning	.54	—											
3	Dial and Table Reading	.51	.65	—										
4	Numerical Operations II	.41	.62	.64	—									
5	Aviation Information	.66	.48	.50	.34	—								
6	Background for Current Affairs	.71	.53	.50	.42	.69	—							
7	Electrical Information	.58	.48	.52	.28	.64	.56	—						
8	Mechanical Principles	.37	.40	.45	.26	.46	.36	.57	—					
9	General Mechanics	.50	.42	.42	.23	.54	.40	.68	.54	—				
10	Tool Functions	.26	.27	.28	.11	.40	.24	.52	.48	.68	—			
11	Speed of Identification	.28	.30	.51	.34	.34	.33	.34	.34	.20	.20	—		
12	Memory for Landmarks	.36	.39	.46	.32	.32	.34	.34	.28	.22	.12	.32	—	
13	Pattern Comprehension	.40	.46	.57	.34	.46	.40	.48	.45	.36	.35	.47	.40	—
14	Colored Form-Word Association	.22	.22	.26	.16	.19	.20	.19	.17	.12	.07	.14	.18	.21
15	Color Recall	.05	.03	.13	.04	.06	.05	.09	.02	.01	.06	.12	.48	.08
16	Word Association	.26	.27	.34	.24	.21	.22	.28	.20	.20	.14	.20	.26	.27
17	Picture Detail Recall	.22	.17	.28	.14	.29	.22	.28	.26	.24	.21	.20	.36	.31
18	Position Recall I	.22	.23	.37	.25	.25	.23	.26	.24	.16	.15	.22	.18	.17
19	Position Recall II	.10	.14	.20	.17	.14	.17	.14	.13	.13	.10	.29	.38	.25
20	Sequence Memory	.32	.27	.29	.25	.25	.30	.28	.12	.18	.06	.18	.26	.28
21	Pattern-Word Association	.23	.25	.31	.20	.22	.23	.26	.22	.17	.13	.24	.46	.36
22	Space Memory	.28	.32	.46	.30	.32	.28	.31	.28	.20	.17	.32	.40	.20
23	Position Memory	.21	.23	.30	.25	.23	.18	.15	.17	.10	.09	.23	.49	.18
24	Number-Word Association	.23	.22	.29	.24	.21	.21	.19	.15	.12	.08	.15	.40	.27
25	Number Association	.14	.27	.31	.34	.08	.14	.07	.06	-.03	-.07	.15	.58	.35
26	Turkish-English Association	.42	.36	.38	.33	.32	.38	.32	.19	.22	.08	.24	.41	.13
27	Pattern Detail Memory	.25	.27	.35	.21	.28	.23	.23	.13	.11	.11	.28	.17	.17
28	Position in Succession	.07	.07	.15	.10	.10	.10	.10	.06	.06	.08	.10	.18	.02
29	Color Memory	.09	.13	.20	.11	.16	.14	.20	.16	.12	.16	.13	.09	.20
30	Age in years	.22	.15	.11	.13	.25	.28	.19	.11	.23	.13	-.01	.34	
31	Education in years	.50	.46	.48	.44	.42	.49	.40	.27	.29	.11	.27		

Note.—Decimal points have been omitted.

which would allow the order of the test administration to be varied. For reasons previously stated, the Color Discrimination Test had to be administered as the first test in the battery. The Color Recall, Position Recall I and Time Sequence tests were designed to be administered immediately after three other tests upon which they depended for subject matter. The Position Recall and Number-Word Associations tests had to be scheduled in the morning sessions in order to allow as much time as possible to elapse before administration of the Picture Detail Recall and Position Memory tests which were included as measures of delayed recall.

Description of the sample. The final administration sample was composed of 718 new Air Force enlisted personnel who were in the early weeks of basic indoctrination training at Lackland Air Force Base, San Antonio, Texas. The testing occurred during the months of October, November, and December, 1953. The sample has a Mean educational level of 11.3 years with a Standard Deviation of 1.88 years. The age distribution for the final sample granted a Mean of 19.35 years with a Standard Deviation of 2.32 years. The geographical origin of 61 cases

was unknown. Of those remaining, approximately 216 came from the Southwestern, 278 from the Southeastern, 98 from the Midwestern and Northcentral, 64 from the Mideastern, and 1 from the Western part of the United States. The final sample was limited to Category I, II, and III airmen⁴ as determined by scores obtained on the Armed Forces Qualification Test. This test is primarily an intellectual measure although it contains spatial and mechanical subtests. Its main function is to eliminate from acceptance into the Armed Forces those individuals who are aptitudinally incapable of making a positive contribution. It has a secondary function of furnishing scores to be used in apportioning talent to the individual armed services. At the time the memory battery was administered, Category IV airmen were being accepted. However, some of these individuals had very little schooling and experienced difficulty in reading and comprehending written instructions.

⁴Categories I, II, and III take in approximately the upper 60% of the draft eligible population.

MEMORY BATTERY AND THE AIRMAN CLASSIFICATION BATTERY

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	M	SD
22	05	26	22	22	10	32	23	28	21	23	14	42	25	07	09	22	50	5.84	1.85
22	03	27	22	22	14	27	25	32	23	22	27	36	27	07	13	15	46	5.87	1.99
26	13	34	28	37	20	29	31	46	30	29	31	38	35	15	20	11	48	5.14	1.89
16	04	24	14	25	17	25	20	30	25	24	34	33	21	10	11	13	44	5.28	1.65
19	06	21	29	25	14	25	22	32	23	21	08	32	28	10	16	25	42	5.22	1.91
20	05	22	22	23	17	30	23	28	18	21	14	38	23	10	14	28	49	5.58	1.78
19	09	28	28	26	14	28	26	31	15	19	07	32	32	10	20	19	40	6.08	1.79
17	02	20	26	24	13	12	22	28	17	15	06	19	23	06	16	11	27	6.00	1.87
12	01	20	24	16	13	18	17	20	10	12	-03	22	13	06	12	23	29	6.40	1.89
07	06	14	21	15	10	06	13	17	09	08	-07	08	11	08	16	13	11	6.23	1.91
14	12	18	20	29	22	18	24	32	23	15	15	24	28	10	13	-01	27	5.14	1.83
42	18	48	26	36	18	26	46	49	40	49	40	58	41	17	18	09	34	4.86	1.98
21	08	27	32	31	17	25	28	36	24	20	18	27	35	13	17	-02	26	5.91	1.93
14	14	37	14	28	13	20	30	36	28	34	34	44	35	09	18	10	19	5.03	1.92
37	17	17	23	23	15	00	12	24	21	25	15	21	14	17	33	-01	05	5.02	2.04
14	23	24	24	39	20	18	40	47	35	50	44	59	36	24	28	04	21	5.11	2.15
28	23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
13	23	39	47	47	27	08	22	32	30	30	16	28	25	15	26	00	16	5.02	1.92
20	15	20	27	—	39	18	26	50	46	48	35	42	32	18	23	09	19	5.01	1.97
30	00	18	08	18	16	15	15	22	22	22	17	20	14	18	12	05	11	5.10	1.95
36	12	40	22	26	15	—	—	36	27	30	36	47	34	20	12	06	08	4.98	1.96
28	24	47	32	20	15	—	—	—	47	43	44	48	41	20	24	08	22	5.12	1.94
34	21	35	30	50	22	15	36	—	47	—	62	38	45	28	16	19	-02	4.95	1.95
34	25	50	30	46	22	18	27	47	—	—	44	57	32	22	23	06	21	5.01	1.92
44	15	44	30	48	22	17	30	43	62	—	—	48	31	13	10	03	22	5.17	2.22
35	14	59	28	42	20	27	47	48	45	57	48	—	42	17	20	16	36	5.14	1.98
09	17	36	25	32	14	20	34	41	28	32	31	42	—	09	15	00	20	5.08	1.94
18	33	24	15	18	18	12	07	20	16	22	13	17	09	—	27	-02	09	4.99	2.08
10	28	26	23	12	06	11	24	19	23	10	20	15	27	—	—	-06	06	5.01	1.93
19	-01	04	00	09	05	08	13	13	-02	06	03	16	00	-02	-06	—	27	5.04	2.06
	05	21	16	19	11	27	24	22	18	21	22	36	20	09	06	27	—	19.84	2.32
																		11.30	1.88

TABLE
CENTROID FACTOR

No.	Variable	I	II	III	IV
1	Word Knowledge	645	-396	-178	-113
2	Arithmetic Reasoning	642	-335	-229	185
3	Dial and Table Reading	746	-246	-068	320
4	Numerical Operations II	553	-203	-283	399
5	Aviation Information	658	-404	063	-146
6	Background for Current Affairs	640	-385	-140	-042
7	Electrical Information	670	-410	206	-233
8	Mechanical Principles	537	-313	261	-101
9	General Mechanics	532	-465	272	-313
10	Tool Functions	403	-337	449	-251
11	Speed of Identification	496	-117	110	235
12	Memory for Landmarks	665	209	-188	-089
13	Pattern Comprehension	609	-181	151	129
14	Colored Form-Word Association	454	243	-162	-163
15	Color Recall	252	303	212	088
16	Word Association	579	358	-080	-152
17	Picture Detail Recall	463	165	317	072
18	Position Recall I	578	319	177	151
19	Position Recall II	341	149	167	174
20	Sequence Memory	373	-078	-160	040
21	Pattern-Word Association	495	180	-128	-154
22	Space Memory	641	282	032	030
23	Position Memory	524	347	-031	093
24	Number-Word Association	567	454	-116	-083
25	Number Association	449	375	-307	100
26	Turkish-English Association	677	306	-296	-206
27	Pattern Detail Memory	505	169	-037	-053
28	Position in Succession	254	179	151	114
29	Color Memory	331	205	281	064
30	Age in years	196	-128	-138	-190
31	Education in years	529	-247	-301	069
	a^2/k	.285	.086	.043	.029

Note.—Decimal points have been omitted.

It was felt that inclusion of these cases in the final sample might introduce unwanted error variance into the test scores.

A small proportion of the final sample were Negroes. No attempt was made to separate these individuals from the analysis group or to treat their scores separately on any of the variables. No women were included in the experimental or final administration samples. Each individual had to sign a statement that he was physically well enough to take the memory battery tests. Any person who did not feel that he was physically able to do so was excused. The only other individuals omitted from this study were those who were totally or partially color blind. These people were tested, but their scores were not included in the final analysis since some of the variables in the battery involved recognition of color.

Scoring and Processing

Answer sheets from the final administration

were scored by machine and the resulting distributions were converted into 9-point normalized standard score form. Intercorrelations of the experimental and reference tests are reported, along with the final distribution statistics, in Table 3.

THE FACTOR SOLUTION

The highest entry in each row of the intercorrelation matrix was used as an initial communality estimate. Twelve factors were extracted using Thurstone's Centroid Method. A full residual matrix was computed after the extraction of each centroid, and adjustments of communality residuals were made prior to the extraction of the succeeding factor. The 12 centroids are reported in Table 4 along with final communalities which

4
MATRIX

V	VI	VII	VIII	IX	X	XI	XII	h^2
236	115	071	-.096	-.081	063	-.070	054	719
-.119	-.074	-.091	-.138	-.088	-.036	-.075	118	687
-.197	043	-.024	-.058	-.051	027	047	-.056	777
-.076	-.133	-.135	-.057	-.090	-.038	-.055	-.069	649
200	076	153	-.026	-.080	-.115	-.049	-.021	713
304	150	089	027	-.097	-.072	-.074	-.042	725
013	062	-.037	059	-.029	040	-.058	051	731
-.192	-.086	068	012	025	-.033	049	-.025	519
-.050	-.217	-.192	-.111	050	121	017	064	791
-.170	-.190	-.125	-.113	094	-.045	023	025	646
-.145	162	180	109	087	026	057	-.126	446
-.137	099	068	-.016	082	048	107	-.030	584
-.233	179	197	-.066	087	087	-.032	063	593
-.100	071	-.037	097	058	-.157	031	142	393
148	113	-.140	-.071	-.111	-.123	179	044	329
-.105	050	-.146	-.069	-.043	132	-.160	-.117	591
134	-.071	144	-.030	-.152	178	072	147	474
113	-.208	122	160	-.056	088	-.098	078	613
129	-.107	-.025	204	123	109	-.071	026	300
101	111	-.058	077	225	103	-.132	119	296
-.183	037	032	110	-.068	095	117	-.070	399
-.133	-.025	094	031	-.124	-.081	-.056	-.069	551
084	-.205	231	-.173	153	-.122	-.047	-.050	580
150	-.183	080	-.233	142	-.050	-.045	-.091	697
-.149	-.095	-.026	046	-.079	-.029	-.108	-.043	501
044	037	-.029	-.036	-.054	101	-.013	-.028	702
-.164	166	122	102	-.046	-.020	-.068	098	385
121	107	-.213	-.062	108	018	-.081	-.134	244
082	191	-.233	-.194	-.088	-.134	065	023	400
169	-.166	-.108	217	-.102	-.046	101	-.108	275
173	-.019	-.064	094	-.034	060	187	-.087	527
.023	.017	.016	.012	.009	.008	.008	.007	

were determined by summing the squares of the factor loadings.

Eleven of the factors were rotated graphically. Of these, eight were rotated orthogonally to approximate Thurstone's definition of simple structure (14, pp. 319-346). Three of the factors were considered to be dependency factors having no psychological significance and were rotated in a special manner. These included the age-education doublet and two dependency factors which were defined by tests based on common study materials. For example, in the Number-Word Association Test, examinees were required to associate numbers and words together so that they later could recall

the numbers from seeing the words. In the Position Memory Test, which was given approximately four hours later, examinees were asked to recall the positions of these number-word combinations as they appeared on the original study page of the Number-Word Association Test. Although both of these tests were based upon the same subject matter, the functions they were designed to measure are essentially different. One test was designed to measure associative memory. The other was designed to measure memory for position in space, which is based on visual imagery. Except for the fact that it was based upon the same study materials, there was no reason to believe

TABLE
ROTATED FACTOR

No.	Variable	I	II	III	IV
1	Word Knowledge	221	051	030	198
2	Arithmetic Reasoning	265	125	027	595
3	Dial and Table Readings	230	119	114	571
4	Numerical Operations II	083	152	033	701
5	Aviation Information	343	240	079	073
6	Background for Current Affairs	147	172	057	175
7	Electrical Information	565	101	054	045
8	Mechanical Principles	540	182	018	103
9	General Mechanics	794	-013	025	092
10	Tool Functions	748	119	090	-009
11	Speed of Identification	088	160	027	177
12	Memory for Landmarks	085	150	144	149
13	Pattern Comprehension	236	122	036	189
14	Colored Form-Word Association	057	246	124	040
15	Color Recall	-009	133	521	-013
16	Word Association	149	120	255	106
17	Picture Detail Recall	184	162	294	-002
18	Position Recall I	106	410	174	112
19	Position Recall II	062	210	059	070
20	Sequence Memory	-003	042	-076	133
21	Pattern-Word Association	148	071	077	074
22	Space Memory	134	385	240	165
23	Position Memory	-013	494	177	142
24	Number-Word Association	024	318	244	088
25	Number Association	-086	271	093	315
26	Turkish-English Association	060	142	196	133
27	Pattern Detail Memory	062	223	105	037
28	Position in Succession	016	080	296	047
29	Color Memory	104	100	576	032
30	Age in years	205	059	-087	047
31	Education in years	091	000	-004	347

Note.—Decimal points have been omitted. Factor XII was not rotated. Factors were defined as follows: I, Mechanical Experience; II, Memory for Position in Space; III, Memory for Color; IV, Numerical Facility; V, Verbal Comprehension; VI, Space-Form Perceptual Speed; VII, Position Recall Content Triplet; VIII, Memory for Position in Temporal Succession; IX, Number-Word Content Doublet; X, Associative Memory; XI, Age-Education Doublet; XII, Residual.

that the Position Memory Test would be related to the Number-Word Association Test any more than it would be related to the Word Association Test or the Number Association Test. However, inspection of the intercorrelation matrix will reveal that the Position Memory Test correlated much higher with the Number-Word Association Test than with any other variable in the battery. Similarly, the Number-Word Association Test had its highest correlation with the Position Memory Test.

A situation such as the one described above usually produces a dependency factor which is undesirable but not ne-

cessarily confusing provided it is recognized as such and rotated in a particular manner. Such a factor is routinely extracted from matrixes of intercorrelations which include age and education variables. Education in years necessarily correlates with age in years for the simple reason that a year of education requires a year's time. The dependency in this case is usually strong enough to produce an age-education doublet factor.

The method for handling dependency factors during the rotational process was fairly simple. Consider the plot which is included as Fig. 4. In this figure, Variable 23 represents the Position Memory

5

MATRIX

V	VI	VII	VIII	IX	X	XI	XII	h^2
733	183	020	092	045	192	-.014	054	713
395	213	-.062	068	-.008	164	-.051	118	693
292	472	004	155	033	144	047	-.056	765
254	142	016	109	-.006	132	046	-.069	644
670	263	013	068	037	052	062	-.021	718
750	188	028	131	-.025	099	090	-.042	721
500	296	026	213	-.051	100	046	051	736
209	363	-.017	070	062	023	046	-.025	524
331	081	-.005	146	122	069	-.007	064	803
105	152	-.071	117	113	-.064	-.041	025	654
150	532	038	185	066	034	081	-.126	435
180	359	-.113	102	088	569	082	-.030	596
219	607	008	169	093	104	-.114	063	589
083	150	-.205	094	-.080	457	123	142	410
-.032	057	047	038	033	093	147	144	331
063	137	042	201	-.083	627	-.125	-.117	607
-.093	254	399	054	177	169	058	147	467
019	185	413	205	105	339	113	078	614
-.013	093	257	342	102	132	134	026	295
272	103	-.050	347	042	208	004	119	292
076	282	-.006	006	-.059	504	142	-.070	406
060	323	101	057	-.054	431	029	-.069	567
059	132	061	058	340	384	-.063	-.050	594
100	007	074	106	339	616	-.096	-.091	707
-.040	095	026	045	-.100	547	013	-.043	513
280	127	002	092	008	728	044	-.028	716
119	377	-.001	099	-.127	374	006	098	398
000	014	034	335	049	099	-.043	-.134	242
029	099	-.027	172	000	051	-.001	023	398
267	-.158	010	-.003	-.085	124	308	-.108	281
469	103	-.010	080	075	218	309	-.087	522

Test and Variable 24 is the Number-Word Association Test. The reference axes were rotated so as to minimize the variance of the other variables on this doublet. The rationale was that the factor (Factor IX) represented a common element in two tests due solely to use of the same subject matter, and there was no reason why any other test in the battery should have been loaded on it.

The final rotated factor matrix is presented in Table 5. Factor XII is included in this matrix although it was never introduced into the rotations. Communalities were reproduced from the rotated loadings and can be compared with those reproduced from the centroids reported in Table 4.

DESCRIPTION OF THE FACTORS

A description of the rotated factors is given in this section. In each case, the tests with the highest loadings are listed and an attempt is made to interpret the factors in terms of the original hypotheses. Also included is a discussion of the factors which were hypothesized but not defined.

Factors Identified

Factor I. Mechanical Experience

No.	Variable	Factor Loading
9	General Mechanics	.794
10	Tool Functions	.748
7	Electrical Information	.565
8	Mechanical Principles	.540
5	Aviation Information	.343

This is clearly the mechanical experi-

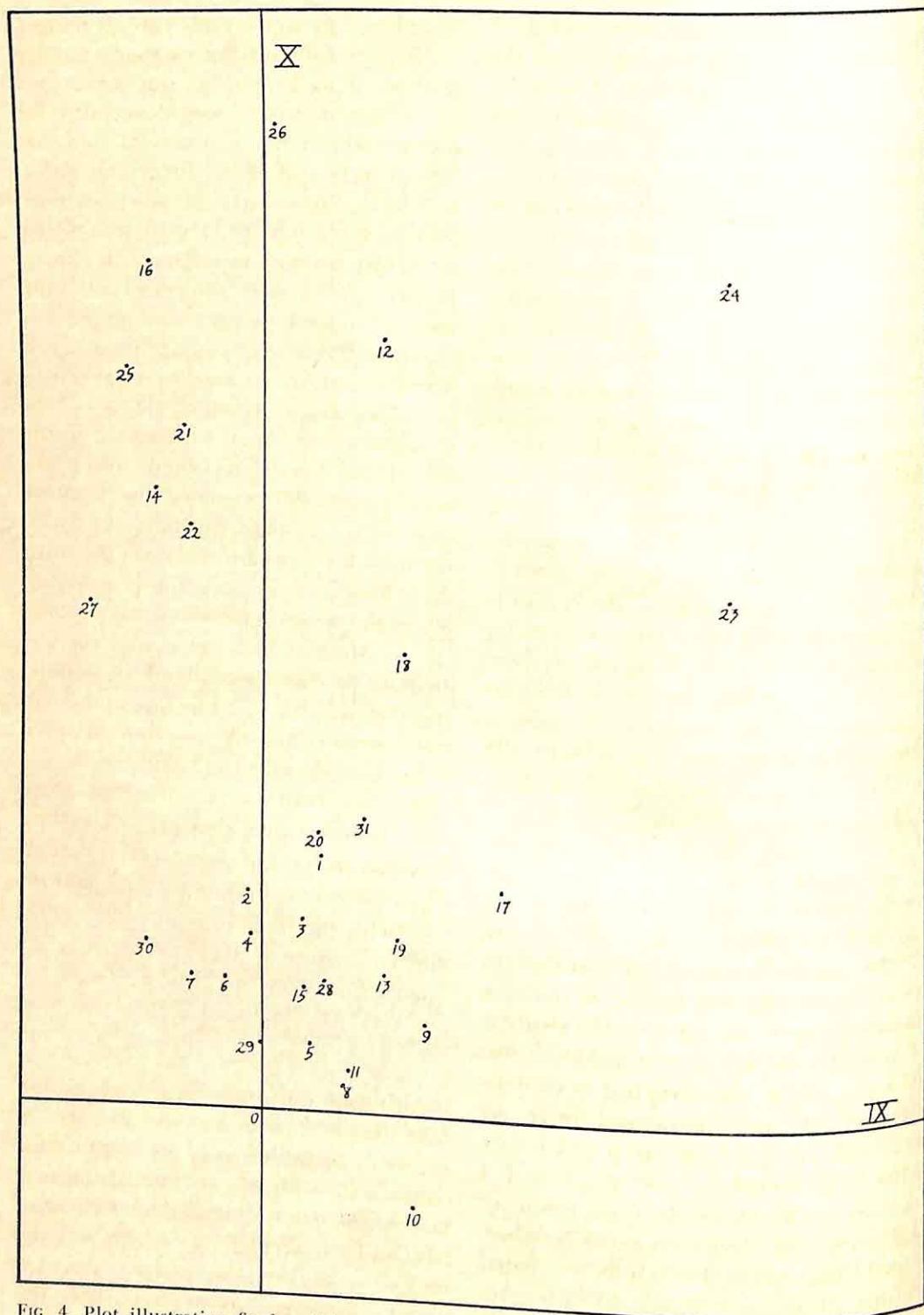


FIG. 4. Plot illustrating final position of reference axes when a factor involving dependency is rotated with a factor not involving dependency.

ence factor which has been found in numerous factor analysis studies of the Airman Classification Battery. It is called an experience factor since tests having the highest loadings on it (General Mechanics, Tool Functions, and Electrical Information) ask questions which can be answered only by examinees who are familiar with common mechanical devices such as automobiles, motorcycles, outboard motors, electrical devices, and so forth. The Mechanical Principles Test is somewhat of an exception since it presents diagrams of operating mechanisms and asks the examinees questions about the movements and functions of their parts. Even here, past experience with machines is beneficial to the examinees and the ones who do best are usually those who have the highest scores on the other mechanical tests. The Aviation Information Test often is found to have a moderate loading on the mechanical experience factor because it contains a number of questions dealing with the mechanical functions of aircraft parts.

Factor II. Memory for Position in Space

No.	Variable	Factor Loading
23	Position Memory	.494
18	Position Recall I	.410
22	Space Memory	.385

This is the memory for position in space factor which was one of the new factors the study was designed to identify. It seems to be adequately defined since all three of the tests designed to contain memory for position in space variance have sizeable loadings on it. The only other test in the battery which had a loading on this factor above the arbitrary .300 level of significance was the Number-Word Association Test. It is not listed as one of the identifying variables because the size of its loading was due to its spurious relationship with the Position Memory Test. Both of these variables

were based upon the same subject matter.

Memory for position in space was hypothesized to be one of the three first-order factors which would account for memory through use of visual imagery. The first two of these, form and color, will be discussed later. Memory for position in space has to do with the ability to picture in one's mind an object in its proper position with respect to other objects called forth by the visual image. For example, if one were attempting to recall where he left his glasses he might "call forth" an image in which he "sees" his glasses on the kitchen window sill to the left of the pot of ivy. Some librarians have remarkable memory for locating books: "It's a small (memory for form) red book (memory for color) on the third shelf between (memory for position in space) the book on physiological psychology by Morgan and Stellar and the elementary psychology textbook by Munn" (associative memory). The fact that a separate memory for the position in space factor was identified indicates that some people are better at recalling the position of objects than they are at recalling the color or form of objects, even though all of these activities might be based on different aspects of the same visual image.

Factor III. Memory for Color

No.	Variable	Factor Loading
29	Color Memory	.576
15	Color Recall	.521

Although only two tests have significant loadings on this factor, its identity is readily apparent. It is the memory for color factor which is another of the new factors this study was designed to identify, and the second one which was designed to account for memory through visual imagery. It appears that the Colored Form-Word Association Test does not measure memory for color in

the same manner as the two tests identifying this factor, since it had a loading of only .124. This bears out the validity of the statements made by examinees interviewed during the experimental administrations of the memory battery who said that they recalled the words and colored forms without resorting to visual imagery. They stated that they relied almost entirely on verbal associations in answering the questions on the recall portion of this test.

Factor IV. Numerical Facility

No.	Variable	Factor Loading
4	Numerical Operations II	.701
2	Arithmetic Reasoning	.594
3	Dial and Table Reading	.571
31	Education in Years	.347
21	Number Association	.315

This is the numerical facility factor which has been identified in all analyses of the Airman Classification Battery. It represents facility in dealing with numbers and is measured in purest form by tests requiring examinees to perform simple numerical operations such as addition, subtraction, multiplication, and division under speeded conditions. However, the factor is more general than this and almost any test involving numbers will measure it to some degree. It is interesting to note that the Number Association Test, which requires the examinees to study and recall pairs of numbers, has a loading above .300.

Factor V. Verbal Comprehension

No.	Variable	Factor Loading
6	Background for Current Affairs	.750
1	Word Knowledge	.733
5	Aviation Information	.670
7	Electrical Information	.500
31	Education in Years	.469
2	Arithmetic Reasoning	.395
9	General Mechanics	.331

This is the very strong verbal comprehension factor which represents the

knowledge of and ability to deal with words. Like the numerical facility factor, it is measured by a variety of tests. In fact, all tests which are presented in a verbal form will be found to have at least a moderate loading on it. Usually the verbal comprehension factor is measured best by a pure vocabulary test, such as Word Knowledge, but occasionally a verbally presented information test, such as Background for Current Affairs, will have a higher loading. The Aviation Information, Arithmetic Reasoning, and General Mechanics tests all consist of verbally presented materials and routinely have appreciable loadings on this factor.

Factor VI. Space-Form-Perceptual Speed (?)

No.	Variable	Factor Loading
13	Pattern Comprehension	.607
11	Speed of Identification	.532
3	Dial and Table Reading	.472
27	Pattern Detail Memory	.377
12	Memory for Landmarks	.359
8	Mechanical Principles	.363
22	Space Memory	.323
7	Electrical Information	.296
21	Pattern-Word Association	.282
18	Position Recall I	.254

The exact nature of this factor is not clear. The Pattern Comprehension, Mechanical Principles, and Electrical Information tests often are found to be loaded on the visualization factor. Speed of Identification and Dial and Table Reading are both perceptual tests. Memory for Landmarks, Pattern Detail Memory, Space Memory, and Pattern-Word Associations were designed to be measures of memory for form.

It has been difficult in past analyses to separate the visualization and perceptual speed factors in airman samples with these tests. It is not known definitely whether this is due to the nature of the tests or of the population. It may be a combination of the two. One thing which

all of the tests loaded on this factor have in common is pictorial items.

Factor VII. Position Recall Content Triplet

No.	Variable	Factor Loading
19	Position Recall II	.413
18	Position Recall I	.399
17	Picture Detail Recall	.257

This factor is due entirely to use of the same subject matter for all three tests. The factor has no psychological significance and will be called simply the position recall content triplet.

Factor VIII. Memory for Position in Temporal Succession

No.	Variable	Factor Loading
20	Sequence Memory	.347
28	Position In Succession	.335
19	Position Recall II	.342

This is the memory for position in temporal succession factor which was one of the new memory factors hypothesized in this study. All of the loadings are small, but this is explained in part by the unreliability of the tests. Strength is lent to acceptance of the factor by virtue of the fact that all three of the tests designed to measure memory for position in temporal succession, and no other tests, have appreciable loadings.

Factor IX. Number-Word Content Doublet

No.	Variable	Factor Loading
23	Position Memory	.340
24	Number-Word Association	.339

This is the number-word content doublet which is due to the dependency of the two tests on the same subject matter. This factor has no psychological significance.

Factor X. Associative Memory

No.	Variable	Factor Loading
26	Turkish-English Association	.728
16	Word Association	.627
24	Number-Word Association	.616
12	Memory for Landmarks	.569
25	Number Association	.547
21	Pattern-Word Association	.504

No.	Variable	Factor Loading
14	Colored Form-Word Association	.457
22	Space Memory	.431
23	Position Memory	.384
27	Pattern Detail Memory	.374
18	Position Recall I	.339

This is the frequently obtained associative memory factor. It has been identified in many previous analyses although it has often been called by other names such as paired-associates memory, rote memory, or simply memory. As has been found in other analyses, paired-associates type tests have the highest loadings. In fact, the seven tests with highest loadings are all of the paired-associates type.

Evidently it is difficult to keep associative memory variance out of any memory test. It may be that many examinees use verbal "crutches" when ever possible to help them memorize facts and events—especially when they recognize that they are going to be called on to recall these phenomena.

It is interesting to note that all of the memory tests not involving incidental recall had significant loadings on the associative memory factor, while none of the incidental memory measures had significant loadings on this factor except for the Position Memory Test. The Position Memory Test is spuriously related to the Number-Word Association Test.

Factor XI. Age-Education Doublet

No.	Variable	Factor Loading
31	Education in Years	.309
32	Age in Years	.308

This is the age-education doublet which has been identified many times before. It has no special significance except that a year of education requires a year's time.

Factors Hypothesized But Not Identified
Incidental Memory. A total of six tests

included in the present battery were designed to measure incidental memory to some degree. A determined effort was made to see that the shortcomings of past attempts to measure this variance were eliminated from the study. For example, the tests were shown to be relatively reliable, the examinees were given ample opportunity to observe materials on which they later were to be tested, and an effort was made to see that the tests measured materials truly encountered incidentally by the examinees. In spite of these precautions there was no factor of incidental memory defined by the study.

Delayed Recall. No delayed recall factor was extracted from the present study although three measures of this hypothesized ability were included in the analysis. Only one of the tests involved a delay of more than five hours between the study and recall periods. For this reason, it can be stated only that, if such an ability exists, it does not begin to operate to a significant extent during the first five hours after exposure to the study materials.

Memory for Form. No memory for form factor was identified by the present study although most of the tests designed to measure this type of variance had significant loadings on Factor VI. Further research is needed to determine whether a memory for form factor exists independent of visualization and perceptual speed.

Visual Imagery. It was hypothesized that three first order factors would be identified in the visual memory area: memory for form, memory for color, and memory for position in space. It was felt that these three types of variance would completely describe the visual image of a past experience and would be sufficiently intercorrelated to produce

a second-order visual imagery factor. Since a memory for form factor was not clearly identified in the study, an oblique solution was not accomplished. The decision was made to await further research before attempting a second-order solution.

Content Factors. Memory for color was the only content factor identified. There was plenty of opportunity for a memory for words factor to be identified, but it failed to appear. Tests designed to measure memory for form did not define a factor, but tended to cluster on a single heterogeneous space-perceptual speed factor. The two tests designed to measure memory for numbers did not define a factor, but this may have been due to under determination. The present study does not settle the question of content factors.

DISCUSSION OF RESULTS

Incidental Memory

Perhaps the most important aspect of the present study was the failure to identify an incidental memory factor. Several carefully prepared tests were included in the analysis which were reliable enough to have produced such a factor if it existed. Special precautions were taken to keep the examinees from suspecting the true nature of the tests. Interviews with the subjects indicated that these precautions were adequate in most instances. It appears unlikely that incidental memory requires anything different from ability to intentionally remember things and events.

Visual Imagery

It is unfortunate that the question of a visual imagery factor was not clearly settled. It seems reasonable that such a factor should exist. The present study defined fairly well a memory for color

factor which corresponds to the ability to remember color through imagery. At least, the examinees reported use of imagery in answering the questions on the two memory for color tests which defined this factor. It was especially interesting to note the strong tendency of examinees to resort to use of verbal associations in learning and remembering the colors of the forms in the Colored Word-Form Association Test. This test had a high loading on the associative memory factor, but almost no loading at all on the memory for color factor. The fact that examinees stated they did not use imagery in recalling these colors was verified by the factor solution.

A position in space factor was identified, although every test which defined it also had a high loading on the associative memory factor. The tests designed to measure memory for form did not, in and of themselves, define a factor, but all of them had substantial loadings on a space visualization factor. Considering all of these findings together, the present study is taken as evidence for special abilities in the area of memory through use of visual imagery. A general visual imagery factor can perhaps be identified in the second order domain. The present study points to the necessity of ruling out examinee use of verbal "memory crutches" if a visual imagery factor is to be defined which is orthogonal to the associative memory factor.

Associative Memory

It was pointed out in the first section of this paper that associative memory is the only well-established memory factor. It has been identified in many studies by independent investigators using a variety of tests and samples. Unfortunately, however, this strong and well-defined factor does not correlate very well with criteria

which most people would agree are highly saturated with memory variance. For example, associative memory contributes so little unique variance to the prediction of success in technical schools that no test of this factor is to be found in the currently operational Air Force selection or classification batteries. Evidently there has been a gross overestimation of the importance of memory in academic success or else the associative memory factor does not measure memory in the same sense as defined by individuals who believe in its importance. Another strange phenomena has to do with the types of tests which do or do not load on this factor. For example, information tests like the Air Force Background for Current Affairs and Aviation Information tests have approximately zero loadings on this factor (.099 and .052 in the present study). It is difficult to understand why recalling the answers to questions about current affairs or aviation should not require an important type of memory. Yet in a factor solution these tests are found to be primarily measures of the verbal factor (.750 and .670 in the present study). The answer seems obvious. The verbal factor is a type of memory factor. The verbal factor in its purest form is reflected by the ability of individuals to recall the meanings of words, and there is no reason to believe that this is essentially different from the ability to recall any other type of verbal material.

In a somewhat broader sense, other factors represent memory abilities. For example, the Numerical Operations Test, which is one of the purest measures of the numerical factor, is little more than rapid recall of the sums, differences, quotients, and products of numbers.

If common factors such as the verbal and numerical factors represent impor-

tant memory abilities, the question naturally arises as to what the so-called associative memory factor measures. Two alternative hypotheses are offered. The first is that associative memory is a rather limited and artificial factor which reflects the extent to which examinees resort to use of "memory crutches" in a formal testing situation. This hypothesis is supported to some extent by the current study in that when special efforts were made to prevent the examinees from using crutches by speeding the presentation of materials or by hiding the purpose of the tests, the resulting scores had little or no loading on the associative memory factor. On the other hand, when no efforts were made to prevent the examinees from using crutches, the tests invariably had high loadings on the associative memory factor.

If one accepts the above hypothesis, it is easy to understand why tests measuring associative memory are not highly predictive of success in meaningful situations. One would expect them to be predictive only to the extent that the criteria in these situations also reflect this type of learning through use of associative crutches.

An alternative hypothesis concerning the nature of the associative memory factor is that it represents ability to reproduce stimuli after limited exposure. This ability may have little relationship to the ability to reproduce the same materials after they have been well integrated through many exposures under various conditions and various contexts. For example, suppose one were to select a group of Turkish words and their Turkish synonyms and assemble them into a paired-associates test. If a group of Americans with no knowledge of the Turkish language were to associate these words in a limited amount of time and

later were required to identify each word with the one with which it was associated, the resulting scores would have a high loading on the associative memory factor, but a low loading on the verbal factor. If the recall portion of this test were administered to a group of Turks who have had many exposures to the pairs in their day to day living, the scores would just as assuredly measure the verbal factor for that sample and would have low loadings on the associative memory factor. In the numerical area, it is entirely possible that a speeded multiplication test would measure associative memory in a fourth grade class which is in the process of memorizing the multiplication tables. The same test would be primarily a measure of the numerical factor if it were to be administered to this same group several years later.

Perhaps the verbal factor represents the ability of the individuals to profit from reinforcements (or to integrate materials) in the verbal area, while the numerical facility factor represents the ability of individuals to profit from reinforcements in the numerical area. An equally tenable position is that the ability to integrate or profit from reinforcement corresponds to a "g" factor and that group factors are the result of differential amounts of experience in particular areas such as verbal and numerical.

Memory for Position in Space

The memory for position in space factor was defined by all three of the tests included in the study to measure this type of variance. As mentioned previously, these tests also had a high loading on the associative memory factor. An attempt should be made to reduce associative memory variance in memory for

position in space tests by speeding the presentation of the stimulus material or by designing the tests in a manner such that the examinees do not feel the need to use verbal associations. Attention also should be called to the close similarity between the three tests defining this factor in the present study. In all cases the examinees were asked to recall the position of materials on a printed page. Additional study is desirable to determine whether this factor can be generalized to include memory for position in three-dimensional space.

SUMMARY AND CONCLUSIONS

The primary purpose of this study was to investigate the factorial structure of the memory area. Nine memory factors were postulated and can be subsumed under five major headings: visual memory, incidental memory, delayed recall, memory for position in temporal succession, and memory content factors. A total of 17 experimental memory tests were constructed, administered, revised, and finally given, along with 14 reference tests, to 718 Air Force enlisted personnel in the early weeks of their basic indoctrination training at Lackland Air Force Base, Texas. The resulting test scores were intercorrelated and a total of 12 centroid factors were extracted from the intercorrelation matrix. Eleven of these factors were rotated orthogonally. Of these, four were identified as reference factors, three as dependency factors having no psychological significance, and four as memory factors. No memory factor was identified that was not originally postulated, although several of the postulated memory factors were not identified.

It is difficult to draw conclusions from a single factor analysis study in a new area since a satisfactory formula for determining the standard error of a factor

loading has not been developed. Factors and factor loadings may vary considerably from population to population and, to some extent, in studies utilizing successive samples drawn from the same population. They are influenced by the factor technique used, by the method of estimating communalities, by the number and composition of reference tests used, by the number of factors extracted, by the method of rotations, and by the idiosyncracies of the individual performing the rotations. For these reasons it is difficult to consider one study as proof for the existence or nonexistence of a factor. When a factor is found over and over again by many analysts using different techniques, samples, and tests, these successive reinforcements give one more confidence that a factor is meaningful. When the results of these analyses are used for reconstructing test batteries which are demonstrated by conventional regression techniques to be more efficient in counseling students into the fields for which they are most qualified, or for better classifying personnel into jobs, then the contributions of such analyses are verified. Subject to the above qualifications, it is believed that the following conclusions are warranted from the results of the present study:

1. If there is a differential loss in retention among individuals extending over more than a single test, it does not occur to a great extent during the first few hours after exposure to stimulus materials.
2. Memory for color represents a special ability which is relatively independent of associative memory.
3. Recalling things and events encountered incidentally requires the same ability as recalling things and events purposefully remembered.
4. There is a special ability for re-

membering the position of objects in space. The present study indicates that this ability is somewhat related to the ability to recall color and more highly related to associative memory. Its relationship to color recall can be explained by the fact that both functions require the production of a visual image. It is probable that much of its relationship to

associative memory in the present study was due to inadequate care in test construction.

5. There is a special ability for remembering the relative position of events in a time series. Although good associative memory may help in this task, it alone cannot account for the superiority of individuals in performing this task.

REFERENCES

1. ACHILLES, E. M. Experimental studies in recall and recognition. *Arch. Psychol.*, 1920, No. 44.
2. ANASTASI, A. Further studies on the memory factor. *Arch. Psychol.*, 1932, No. 142.
3. CARLSON, H. S. Factor analysis of memory ability. *J. exp. Psychol.*, 1937, 63, 477-492.
4. EBBINGHAUS, H. *Memory. A contribution to experimental psychology.* (Trans. by H. A. Ruger.) New York: Columbia Univer. Press, 1913.
5. FORLANDO, G. & BUNSCHWIG, L. Perseveration in relation to reminiscence and recall. *Brit. J. educ. Psychol.*, 1938, 8, 178-187.
6. FRENCH, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometr. Monogr.*, 1951, No. 5.
7. GRANT, M. E. Some theories and experiments in the field of memory. *J. educ. Psychol.* 1932, 23, 511-527.
8. GUILFORD, J. P. (Ed.) *Printed classification tests.* AAF Aviation Psychol. Prog. Res. Rep. No. 5. Washington, D.C. U.S. Government Printing Office, 1947.
9. HUMPHREYS, L. G. & FRUCHTER, B. Re-analysis of Memory Batteries I and II. *San Antonio Psychological Research Unit Research Bulletin T45-15*, 1945.
10. KELLY, PAUL H. A factor analysis of memory ability. Unpublished doctoral dissertation, Princeton Univer., 1954.
11. SHELLOW, SADIE M. Individual differences in incidental memory. *Arch. Psychol.*, 1923, No. 64.
12. THURSTONE, L. L. Primary mental abilities. *Psychometr. Monogr.*, 1938, No. 1.
13. THURSTONE, L. L. Current issues in factor analysis. *Psychol. Bull.*, 1940, 37, 189-236.
14. THURSTONE, L. L. *Multiple-factor analysis.* Chicago Univer. Chicago Press, 1940.
15. THURSTONE, L. L. Theories of intelligence. *Sci. Mon.*, 1946, 62, 101-112.
16. VERNON, P. E. *The measurement of abilities.* London: Univer. London Press, 1940.

Psychological Monographs: General and Applied

The Effect of a Disturbing Situation Upon the Stability of Various Projective Tests¹

RHODA LEE FISHER

Baylor University College of Medicine, Houston, Texas

THE QUESTION whether projective tests are stable has been investigated in terms of a whole gamut of techniques and procedures (1, 2, 3, 15, 20, 30, 32, 33, 34, 35, 36, 41, 42, 46, 47, 49, 51, 53, 54, 48, 21, 44, 11, 50, 38). Results of such investigations have been varied and somewhat conflictual. One can find studies (19, 40, 23, 25) which demonstrate high stability of projective tests and others (7, 6, 55, 10) which indicate just the opposite to be true. The conflicting results in this area that one finds in the literature seem to be due to a number of factors.

To begin with there is little agreement as to what experimental conditions constitute a fair evaluation of the stability of the tests. Should one measure the stability of a test by test-retest with an intervening time period of brief duration? Or should the intervening period be of long duration? Should stability be measured by exposing the subject to conditions of special instruction or stress and noting if they produce changes in the subject's responses? If so, what kinds of instructions or what forms of stress should be used? In the literature one finds that

"special experimental conditions" designed to produce changes in test responses have varied considerably. They have included intellectualized requests to take a special attitude toward the test and even the hypnotic induction of special moods.

A second difficulty that has arisen in evaluating the problem of change in projective tests has revolved about the fact that there are no standards for judging what shifts in test responses may be interpreted as representing real or significant shifts in the individual from whom they have been obtained. Is a just statistically significant increase in the number of responses on the Rorschach an indication of a psychologically meaningful shift? Should one interpret an obvious increase in a size of a figure drawing upon retest as a psychologically significant shift?

A third point which has complicated the evaluation of projective test stability has to do with the level of the personality trait involved. That is, one can get significant shifts in test materials which can be interpreted as personality shifts. But in one instance the investigator may interpret the shift as having occurred in a personality area so peripheral as to be meaningless and unimportant. Another investigator may consider the same area of personality shift to be much more important.

¹This monograph is based on a thesis submitted to the University of Chicago in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

The writer wishes to express her indebtedness to William Henry, Hedda Bolgar, and Seymour Fisher for their invaluable help during the planning and execution of this study.

A fourth difficulty has to do with the lumping of all projective tests together and speaking of them in a stereotyped oversimplified way. One forgets that there is a range of projective tests, each with perhaps its own conditions of stability. This whole point is now vague and poorly defined because no one has studied the interrelationships of changes occurring in a variety of projective tests given to each subject under specific conditions.

PURPOSE

The present study aimed to get at the issue of projective test stability, but in such a manner as to avoid some of the difficulties involved in previous work in the area. Specifically, the study aimed to deal in an exploratory fashion with several issues.

1. One aim was to determine to what degree a psychological stress growing out of a real life situation affects the general pattern of response of subjects to a variety of projective tests. It was the intent to explore subjects' patterns of responses in such stress situations, not in terms of single indices, but rather by means of a variety of indices ranging from the simple to the complex. Thus, the possible effects of the stress situation could be scrutinized in terms of a great many different kinds of cues.

2. A second purpose was to determine the differential sensitivity to situational stress of the various indices derived from each test. That is, the objective was to evaluate the relative sensitivity and insensitivity of the various systems of measurement that may be applied to projective tests.

3. A third goal was to determine whether there is a hierarchy of sensitivity of the various projective tests to situational stress. Does one test consistently

show more change under stress than the other tests over the whole range of subjects? Does one find more subtle hierarchical patterns such that, if one test shifts under stress, another given test also consistently shifts, but a third test consistently does not shift?

Over-all, then, the purpose of this study was to find out whether a stress situation can produce significant shifts in projective test responses; whether there are inferior or superior ways for detecting these shifts; whether individual projective tests manifest significant differences in their susceptibility to shift; and whether there are characteristic patterns of shift.

PROCEDURE

The general plan of the project was as follows. A battery of tests was administered to an *experimental* group immediately following a disturbing physical examination and then the same subjects were retested after a period of five days. The battery of tests was administered to another group on two occasions with the same five day period intervening, but a disturbing examination did not precede the first administration of the battery. This second group served a control function. The changes occurring in this second group simply with the passage of time could be contrasted to the changes occurring in the first group where the test-retest conditions differed rather markedly in terms of amount of situational anxiety present.

In order to study realistically the impact of temporary anxiety on responses to projective tests, it was necessary to make sure that the "cards were not stacked" in advance against positive findings. That is, if one were to set up an experimental situation in which local temporary anxiety was defined in terms

of some artificial laboratory stress, with but the most peripheral significance to the personality, one would be expecting too much from the sensitivity of projective tests. Consequently, a real life source of anxiety was selected. The opportunity to make use of such a source presented itself in certain aspects of the admission procedure to which patients are subject in a psychiatric state hospital setup. Each female patient who is admitted to Elgin State Hospital undergoes a gynecological examination at some time during the first week or so after she is admitted. Repeated observation has indicated that this examination is very traumatic to most of the patients. Frequently it is necessary to force patients to the examination and the whole procedure may be accompanied by crying and screaming and protests against being subjected to such "indignity." Some patients submit to the examination only after being drugged, and it is not uncommon to have to restrain patients forcefully in order to keep them on the examination table. Attendants and nurses frequently note that following the examination patients are tremulous, unusually frightened, and occasionally faint.

The staff psychiatrists on the admission wards all agree that, of all the admission procedures, the gynecological examination is the most disconcerting to patients. In general, it is clear that the average female patient who comes to a state hospital with some psychic disturbance finds it humiliating and provocative of much anxiety to expose her genitals to a male physician. Indeed, it is widely recognized by gynecologists that a vaginal examination provokes considerable anxiety in patients. Haas has intensely studied the meaning of the vaginal examination and analysed the various sorts of anxieties activated by the procedure. Haas as quoted in Bellak (5, p. 91) indicates that "... adult women react to a gynecologic examination with anxiety, as to a threat of being injured. Fears are centered around the genitals to a much higher degree than any other part of the body. Moreover, to the adult woman the first

vaginal examination may also have connotations, since it is the first time she exposes herself to anyone but her husband. In spite of the professional setting, the experience may have a sexual meaning to the patient and, therefore, have the meaning of the forbidden."

It was considered, then, that the gynecological examination constituted for most of the patients a source of considerable embarrassment and created a period of considerable stress. One cannot really define with any precision what aspects of the examination are most threatening. There are obviously elements of humiliation, sexual temptation, and sexual embarrassment in the procedure. At the same time, there is also present the literal threat of physical injury. The gynecologist displays a variety of instruments and in the course of his examination inflicts a fair amount of pain on the subject. Finally, there is a threat posed in the sense that the patient is forced to submit, forced to assume a position of vulnerable passivity. Even though one cannot define which of these elements of threat is of relatively greatest prominence to each patient, one can say that they constitute a gross constellation of great threat. What perhaps should be most emphasized is that the gynecological stress situation is *relatively* much more real, ego involving, and personal than the typical stress which has been used in past research concerning the stability of projective tests. Most typically, the stress situations in past research in this area have involved such things as watching a battle movie, being observed by an audience, and being exposed to an examiner who attempts to assume a hostile attitude. All of these attempts to produce stress have an aura of artificiality about them. It was this artificiality and unrealness which was avoided in the present study.

Of course, the meaning of the gynecological examination to any one individual patient could not be determined in the present experimental design. It is possible that some patients were not significantly disturbed by the examination. However, this is a defect which is inherent in all studies involving an evaluation of the effects of stress or anxiety upon performance. The existence of a stress effect must be conceptualized out of numerous observations concerning the overt behavioral disturbance shown by many patients in the course of being examined.

In the present study, the control and experimental groups each comprised 25 cases of first admission women patients. Of these 50 patients, 38 had been diagnosed by the psychiatry staff as being psychotic and 12 had been diagnosed as being without psychosis. Within the experimental group, 68% were diagnosed as psychotic and 32% as being without psychosis. Only co-operative patients were included for it was felt that shifts in projective material upon retest might otherwise occur as the result of such extraneous factors as fluctuations in attention or willingness to exert effort. All subjects fell within the age range 20 to 40. It was felt to be important that the study be free from the complications of very low intellectual endowment. Education was taken as a rough estimate of intelligence and only persons with the equivalent of a grammar school education were included. This helped to rule out misunderstandings in instructions and kept the experimental and control groups roughly equated intellectually. Only first admission patients were used. This was in order to minimize differentials in anxiety associated with differential familiarity with the hospital environment.

Since the majority of the patients used were psychotic, it may be considered that the evaluation of the stability of the various tests used is more stringent than it would be if all the subjects were normal. That is, it has been clearly demonstrated (22) that psychotic patients tend to be considerably more variable in their retest performances over a wide range of tasks and tests. Reaction time, motor performance, and reactions to many other complex tasks have all been shown to be more variable upon retest in psychotics than in normals. In order to make it possible to check to some degree on the relationship of psychosis to test-retest variability, 32% of the

subjects included in the experimental group were without psychosis and, likewise, 16% of the subjects included in the control group were without psychosis. Thus, the variability manifested by the psychotic patients could be compared with the variability of the nonpsychotics.

The projective tests were always given in the same order, on an individual basis, and by the same individual examiner who did all of the testing. At no time were patients urged or asked probing questions. It was felt that arbitrary questioning or urging would introduce an uncontrolled variable that might cause artificial shifts on retesting. All materials obtained were spontaneous productions. The battery used in the study was chosen so as to include tests which seem to tap different aspects of the personality structure. They are listed below in the order in which they were administered.

1. The Rorschach Inkblot Test.

2. A series of Thematic Apperception cards, including seven pictures from the widely used Murray TAT series and four from the original Worchester series of the Murray cards. Cards were selected which pictured a wide variety of situations, including a number with sexual implications.

3. A widely used variation of the Figure Drawing Test was used. The subject was asked to draw the full length picture of a person and, upon completion of this first drawing, to draw the full length picture of a person of the opposite sex. Finally, the subject was requested to make up a story which would involve the two figures drawn.

4. A word association test was also used. It consisted of an original list of 20 stimulus words (Appendix I). Ten of these words were chosen for their "neutral" quality and the other 10 were chosen for their personal emotional quality (27, 29, 43). Such words as "curtain," "window," and "chair," were considered neutral. Words such as "nipple," "blood," "fear," and "baby," were considered personal. The list was so arranged that these two kinds of words were randomly distributed.

METHOD OF ANALYSIS AND RESULTS

The analysis of the data presented a very difficult problem. All research with projective techniques has been hampered by the difficulty involved in quantifying and making objective projective data. Over and over again one finds this difficulty mentioned in the literature (4, 24, 37). How can one compare the Rorschach pattern of a first testing session with that

of the second testing? What is to be considered a significant change? How does one compare two TATs? If one story in a TAT series changes markedly does this constitute a significant change in total TAT results? How many TAT stories have to change before a significant change in test results can be considered to have occurred?

Four main approaches to the analysis of projective data have in the past been attempted.

1. *Ratings.* This technique usually utilizes trained judges who are asked to read the projective materials and rate specific variables which presumably are measured by the projective data.

2. *Individual Scores.* This method involves use of the customary individual scores which are derived from projective tests. On the Rorschach, this type of score would be exemplified by such variables as number of movement responses, level of form accuracy, and number of color responses. On a word association test, this type of individual score would be exemplified by average reaction time.

3. *Matching Techniques.* The matching technique is one in which the test and retest items are mixed together and judges are asked to match those responses which were given by the same subject. Vernon (52) and Troup (51) and also Holzberg (23) in their work on Rorschach reliability, have used this method.

4. *Weighted Scores.* Most recently, rather complicated techniques for quantifying projective materials have become increasingly prominent. Buhler and Lefever (8) attempted to differentiate different diagnostic groups on the basis of various Rorschach ratios or "clusters." Elizur (14) and Fisher and Hinds (18) have devised weighted scores which aim at quantitative evaluation of the hostility expressed in a Rorschach protocol. Fisher (17), in a study of rigidity phenomena, successfully used weighted scores for measuring maladjustment and rigidity on the Rorschach. These quantifying techniques attempt to duplicate the complexity of appraisal that usually goes into a clinical evaluation of a projective test. Scores are derived from quantitative evaluation of those pattern relationships that are actually considered in a careful clinical evaluation of a record. For example, in analyzing the degree of maladjustment in a Rorschach record, the clinician usually looks for such things as the number of movement responses, level of form quality, and manner in which color is used. Within the clinician's own

personal frame of reference he weights each of these things in terms of his clinical experience. Paralleling this clinical approach, the weighted score technique assigns weights to factors and patterns in proportion to their usual clinical importance in evaluating a given variable.

Of the four methods enumerated above, three were applied to the evaluation of the data obtained in the present study. The way in which each of these methods was applied is described below.

The Rating Approach

The rating method was first applied in a pilot study preceding the present project. In this preliminary study, raters were asked to rate for 10 cases the shifts in projective data for 15 different variables. Raters were told to evaluate the projective materials in terms of the clinical signs, they usually used in analyzing such tests for the given variables. It was almost immediately discovered that the raters found it difficult to make the judgments required. They felt they could not actually discern the fine distinctions demanded by the 15 different rating continua. It was because of such initial rater reactions that the preliminary rating continua were revised before being used in the present study. The new scales involved rating only five variables and these variables were those which the raters had found easiest to evaluate. The five variables chosen for rating were, furthermore, considered to be representative of those areas usually given prominence in careful qualitative analyses of projective data.

Three sets of ratings were obtained. Two raters each studied all of the 50 cases and rated each case. The third "rater" was made up of a composite of five clinical psychologists. Each of these psychologists rated 10 cases. They were asked to rate five cases in the control

group and five cases in the experimental group. The test materials were coded so that the raters could not identify which were obtained from control group subjects and which were obtained from experimental subjects. The necessity for the composite rater arose when it was found to be impossible to find more than two raters sufficiently motivated to rate all 50 cases. All of the raters had had at least two years of experience in the use of projective tests in a clinical setting. They were asked to rate shifts in TAT, Figure Drawing, and Rorschach. The rating continua upon which they made their judgments were set up in such a way that they could indicate intensity shifts for each variable, from first to second testing, in the following terms: "great increase," "moderate increase," "no change," "moderate decrease," or "great decrease."

In dealing with the Rorschach, the raters were asked to evaluate shifts in the following variables: rigidity, maladjustment, energy invested, sexual preoccupation, and hostility. Each of these rating categories is defined below:

1. Rigidity was meant to refer to tightness and constriction. It refers to inability to adjust to changing circumstances and inability to vary one's responses.

2. Maladjustment refers to amount of breakdown in personality controls. To what degree is the individual distorting reality, living in terms of unrealistic fantasy, or resorting to ineffectual compensations to handle severe inner conflicts?

3. Energy invested refers to the degree to which the subject seemed to really put herself into the task (Appendix II). How far did the subject become involved in the task and how openly did she permit herself to deal with the materials? Were reactions superficial and guarded or did the subject participate in a personally involved fashion? How much real effort was put forth in responding to the inkblot stimuli or the TAT pictures.

4. Sexual preoccupation was defined in terms of number of sexually tinged responses occurring in the record. To what degree did the subject

seem to be concerned with sexual ideas, sexual fantasies, and sexual anxieties?

5. Hostility was judged according to the amount of hostility expressed in the record. Raters judged to what extent overt and symbolic forms of hostility response in the Rorschach shifted.

The ratings of the Rorschachs of the control and experimental groups were compared in terms of average amount of change from first to second testing.² The mean group shift in both groups was insignificant. For the most part the rating variables were judged to have remained unchanged. Thus, for both the experimental and control groups, the group means ranged from 2.9 to 3.2. The ratings for each subject, in the main, showed little variation. In the experimental group there were only seven instances in which any subject was judged by any of the raters to have shifted extremely (1 or 5) for any of the variables.

Thematic Apperception Test ratings were carried out in a fashion analogous to the Rorschach ratings. The three raters evaluated TAT shifts in terms of energy invested, hostility directed outward, hostility directed inward, optimism-pessimism, and attitude toward the opposite sex. The definition of "energy invested" supplied to raters when dealing with the TAT materials was identical with that used for the Rorschach ratings. The remaining rating variables were defined as follows:

1. Hostility directed outward refers to the degree to which the TAT stories revolve about themes in which characters are directing hostility toward others in an open fashion.

2. Hostility directed inward refers to the degree to which the TAT themes revolve about characters who turn their hostility inward, blame

² Detailed tables describing the various results obtained from the rating analysis on the data have been deposited with the American Documentation Institute. Order Document No. 5731, remitting \$1.75 for 35-mm. microfilm or \$2.50 for 6 by 8 in. photocopies.

themselves, and get themselves into trouble.

3. The optimism-pessimism variable refers to the increase or decrease in the subject's tendency to see things in a happy, favorable light. Do the stories end happily? Do things turn out well? Or are things in the environment seen as unfavorable and unhappy?

4. Attitude toward men refers to the way in which the individual describes the male figures in the various TAT stories. Does she express relatively more or less positive feeling toward them from the first to the second testing?

Consistent with the findings on the Rorschach test the mean amount of shift for the above variables was small and insignificant in both the control and experimental groups. The greatest average amount of shift in the control group was .1; and in the experimental group, also .1. Only three experimental subjects were judged by any of the raters to have shifted extremely (1 or 5) relative to any of the rating variables. Likewise, only two control group subjects were judged by any of the raters to have shown an extreme shift relative to any of the rating variables.

Raters were asked to rate the figure drawings on the following continua: energy invested, hostility directed outward, hostility directed inward, maladjustment, and attitude toward men. All of the rating variables, with the exception of maladjustment, were defined in the same way as they were for the TAT ratings. Maladjustment was defined in the same way as it was for the Rorschach ratings.

An analysis of the figure drawing ratings indicated that the mean amount of shift for each of the variables from the first to the second testings was insignificant in either the control or experimental groups. The various group means for the two groups were quite similar in that they all clustered in the vicinity of 3 (no change). Thus, for the control group, the group means ranged from

3.0 to 3.1; and for the experimental group, the group means ranged from 3.0 to 3.1.

It is important to note that all the rating procedures which were applied to the various projective tests involve a direct comparison of test responses with retest responses. That is, the raters were asked to express their judgments in terms of over-all units of change rather than to make separate judgments about the test and retest responses and use the difference between test and retest as the unit of change. The raters did not have to make an explicit evaluation of the degree of presence of each separate variable in the test and also in the retest data. One might therefore raise some pertinent questions which ought to be dealt with.

First of all, is it possible that the variables which the judges dealt with in terms of direct units of change cannot actually be judged reliably in terms of their own individual continua? Perhaps using degree of change as the direct unit of change covers over the unreliability of rating of the individual variables. If the individual variables could not be rated reliably, this would of course seriously decrease the meaningfulness of the over-all results obtaining by means of the rating procedure.

A second question that one ought to consider is the possibility that ratings made directly in terms of units of change are particularly susceptible to any bias that raters might have about the stability of projective tests. For example, if a rater felt a need to demonstrate the stability of projective tests, might he not have a tendency to use the category "no change" with unrealistic frequency?

In order to check on these possibilities a segment of the data was reanalyzed by means of a new rating procedure. The

test Rorschachs, test TATs and test Figure drawings of 10 of the experimental subjects were pulled at random from the total experimental group. That is, the test and retest responses of 10 subjects were re-evaluated. All forms of identification were removed from the test materials and there were no clues as to which materials were obtained at the first testing and which were obtained at the second testing. All of the material was then shuffled and presented to two individual raters. These raters were asked to rate each test in terms of the five given continua which had previously been used in terms of degree of change units. For example, raters were asked to evaluate the Rorschachs in terms of degree of rigidity, degree of maladjustment, degree of hostility, degree of sexual preoccupation, and degree of energy invested. All of the variables were defined in exactly the same way as they had been described to previous raters; and all ratings were made on a five point scale varying from a minimum quantity of the variable to a very large amount of the variable. The raters were instructed to use as the middle point of the scale the typical maladjusted patient that one would find in a hospital for patients with chronic psychotic disturbances. Over-all, then, each rater made judgments concerning 60 different tests with each judgment on a five point continuum.

There was marked agreement between the two judges in their ratings of the five Rorschach variables for both the test and retest materials. Considering all the 10 categories of ratings involved, it was found that, for six of the categories, none of the judgments made by the two raters differed by more than one rating unit. For the remaining four categories, 10% of the ratings differed by two rating

units. It was clear from these results that the two raters were able to make their ratings of the Rorschach variables with a substantial and satisfactory degree of agreement.

For the 10 categories of Thematic Apperception Test ratings, there were eight in which disagreements greater than one rating unit did not occur. In two of the categories, 10% of the ratings differed by as much as two rating units, and for one category (energy invested) 10% of the ratings differed as much as three rating units. Here too the degree of agreement between the two raters was very high. Only in the instance of energy invested for the first testing might one raise a question concerning the agreement of the raters, since 10% of the ratings were different by as much as three units. However, when one considers that, for the retest series, the raters did not disagree about any of their judgments as much as two units, the results obtained in the test series may not be considered to have much import.

The ratings for the Figure Drawings given by the two judges showed the highest degree of agreement. There was only one category in which ratings of more than one unit disagreement occur and only 10% of the ratings were in this category. Here too the raters demonstrated that they could evaluate the variables reliably within a similar judgmental framework.

Because the ratings given by the two judges were by and large so much in agreement, they were pooled in the analysis that was made to determine if the retest responses differed from the first test responses. Thus, the average initial ratings and the average retest ratings of the two judges for each of five variables formulated for each were compared. In

the instance of the mean test and retest rating for the five variables, the means did not in any case even reach one rating unit difference. Likewise, the mean of each subject's test-retest shifts did not attain one rating unit. The same "no change" pattern held true for the Thematic Apperception Test. Somewhat more change was found in the comparison of test-retest ratings for the Figure Drawings. None of the mean test and retest differences attained a whole rating unit. In two instances they did attain a half of a rating unit. But none of the differences were statistically significant. It may be concluded on the basis of these results that the variables involved can be reliably rated. Furthermore, it would appear that, even with the elimination of the bias potentially inherent in the method of rating based on the *direct* evaluation of degree of change, the pattern of results still indicated that the various tests did not shift significantly from initial testing to retesting.

Weighted Score Approach

Weighted score evaluations of the following Rorschach variables were utilized: rigidity, hostility, maladjustment, and energy invested. The weighted score techniques for evaluating maladjustment and rigidity in the Rorschach were derived from previous studies carried out by Fisher (17) and Fisher and Hinds (18). In both of these studies it was shown that the maladjustment score and the rigidity score were capable of differentiating significantly among meaningful criteria. Since the appearance of these two studies, other data have been collected which confirm the usefulness of the maladjustment and rigidity indices (16, 26). George De Vos (12) found, in a study of Japanese Americans living in Chicago,

that the Japanese who displayed rigid character traits could be differentiated from a control group of American subjects in terms of the Fisher rigidity score. In still another study Robert Roman (45) demonstrated that the Fisher rigidity score significantly differentiated a group high in ethnocentrism from a group low in ethnocentrism; whereas individual Rorschach scores did not successfully differentiate the groups. The validity of the Fisher Maladjustment score has been reinforced by two recent studies. Thetford and De Vos, as described in De Vos (12), state:

A report by Thetford and De Vos using Fisher's Rigidity and Maladjustment scores on Beck's normal, neurotic and schizophrenic records is now in final preparation. On the Maladjustment score Fisher found means of 36.9, 59.7, and 85.5 respectively for his groups (entirely women) whereas we found means of 33.1, 59.4, and 85.2 for combined men and women. Women alone score 30.0, 52.97, and 93.88 respectively.

A critical cut off 2σ above the mean for the normals includes: normal-4 (7%); neurotic-15 (50%); schizophrenic-24 (80%).

The correspondence between Fisher's results and those of Thetford and De Vos are indeed striking. Fine, Fulkerson, and Phillips (16), using life history data as the criterion, have found that the Fisher Maladjustment score significantly differentiated low attaining normal individuals from high attaining normal individuals. Differences between the groups were significant at the .001 level of confidence. Note the following conclusion that is drawn in the article: "Fisher . . . has developed a scale which seems to be a valid measure of intrapsychic adjustment. Fisher derived this scale on the basis of the literature, taking into account those components which have been said to go with pathological behavior. And both he and Thetford and De Vos found the scale did correlate with clinical

cally determined pathology." All in all, then, it would appear that the use of the Fisher maladjustment and rigidity indices is justified in terms of the early promising results they have shown up to this point.

The scoring method for evaluating energy invested was devised specifically within the context of the present study. It was devised purely on a common sense exploratory basis. The character of the score is more clearly understandable if one examines in Appendix I the variables that contribute to it.

When a comparison was made of the various Rorschach weighted scores obtained for each group on the first and second testings, no significant shifts were found in either group. Differences between the weighted score means, as shown in Table 1, tended to be negligible.

The TAT protocols were scored for hostility, energy invested, attitude toward men, and optimism-pessimism. These categories have the same *general* meaning here as assigned to them on the rating

scales. The specific details regarding how these scores were derived may be found in Fisher and Hinds (18), and in Appendixes III, IV, and V. Comparing the various TAT weighted score means on test-retest (Tables 2 and 3) demonstrated no significant shifts. Differences between test and retest mean scores for optimism-pessimism, attitude toward men, and energy invested range in the experimental group from 0 to 1.2 and in the control group from 0 to 2.1. *T* scores for these groups for each of the mean differences ranged in the experimental group from 0 to 1.0 and in the control group for 0 to .5. As shown in Table 3, test-retest differences in percentages for the various categories of hostility response vary in the experimental group from 0 to .1 and likewise in the control group from 0 to .1. None of these differences proved to be significant.

The Figure Drawing stories were assigned weighted scores in the same way as were the TAT stories. However, the figure drawings themselves had to be handled in a somewhat different fashion

TABLE 1

TEST-RETEST MEANS FOR THE VARIOUS WEIGHTED RORSCHACH SCORES IN THE EXPERIMENTAL AND CONTROL GROUPS AND THE SIGNIFICANCE OF THE TEST-RETEST MEAN DIFFERENCES

DIFFERENCES AND THE SIGNIFICANCE OF THE TEST-RETEST MEAN DIFFERENCES					
	Maladjustment	Rigidity	Energy Invested	Hostility	
				Overt	Symbolic
Experimental Group					
Test	73.8	46.6	30.5	1.4	2.9
Retest	75.6	44.2	28.2	1.6	2.8
Differences	1.8	2.4	2.3	.2	.1
<i>T</i> scores	.2 ^a	.5	.9	0	.1
Control Group					
Test	80.4	47.5	32.9		
Retest	78.8	45.6	32.8	1.6	3.2
Differences	1.6	1.9	.1	1.6	3.7
<i>T</i> scores	.1	.4	0	0	.5
				0	.5

^a These *T* scores were obtained by dividing the differences by the standard error of differences.

^a These *T* scores were obtained by the method referred to in Edwards (13) as the "*T*" to test the difference between correlated means.

TABLE 2

TEST-RETEST MEANS OF WEIGHTED TAT SCORES FOR OPTIMISM-PESSIMISM, ENERGY INVESTED, AND ATTITUDE TOWARD MEN IN BOTH EXPERIMENTAL AND CONTROL GROUPS

	Optimism-Pessimism			Attitude Toward Men			Energy Invested
	Optimistic	Pessimistic	Indefinite	Favorable	Unfavorable	Indefinite	
Experimental Group							
Test	.5	3.7	6.8	1.2	2.9	6.9	35.4
Retest	.5	3.8	6.7	.8	3.1	6.9	34.2
Mean difference	0	.1	.1	.4	.2	.0	1.2
T scores	0	.2	.2	.2	1.0	0	.6
Control Group							
Test	.6	3.6	6.8	1.1	2.5	7.4	36.3
Retest	.6	3.6	7.2	1.2	2.4	7.4	34.2
Mean difference	0	0	.4	.1	.1	0	2.1
T scores	0 ^a	0	.5	.3	.2	0	0

^a T scores were obtained by the method used to test the differences between correlated means.

because of the extreme subjectivity involved in evaluating them. Rather than attempting to evaluate the drawings on the basis of such broad variables as used for the TAT, they were evaluated in much more limited specific terms. A more detailed description of the handling of this material is provided in a later sec-

tion dealing with individual score evaluations of the data. There were no shifts of significance in the Figure Drawing stories. The mean differences for the experimental group ranged from 0 to .4 and for the control group from 0 to .3. The percentage differences in hostility score for the experimental group varied

TABLE 3

PERCENTAGES OF DIFFERENT TYPES OF TAT HOSTILITY RESPONSES ON TEST AND RETEST FOR BOTH THE EXPERIMENTAL AND CONTROL GROUPS

	Openly Expressed Hostility	Indirect or Rationalized Hostility	Ambivalently Expressed Hostility	Neutral Content	Positive or Favorable Feeling Expressed
Experimental Group					
Test	.2	.1	.1	.5	.1
Retest	.2	.1	0	.5	.1
Percentage differences	0 ^a	0	.1	0	0
Control Group					
Test	.1	.2	.1	.5	.1
Retest	.2	.2	.1	.5	.1
Percentage differences	.1	0	0	0	0

^a None of the test-retest differences obtained for the variables described in this table even approaches the 5% level of significance.

TABLE 4
RORSCHACH INDIVIDUAL SCORE TEST-RETEST
MEANS AND DIFFERENCES IN MEANS,
FOR THE EXPERIMENTAL GROUP

Factors	Test	Retest	Differences ^a
<i>W</i>	4.6	4.0	.6
<i>D</i>	8.1	9.6	1.5
<i>Dd</i>	2.4	3.3	.9
<i>M</i>	1.0	1.1	.1
<i>C</i>	.2	.4	.2
<i>CF</i>	.8	.6	.2
<i>FC</i>	.9	1.1	.2
<i>FY</i>	2.6	2.8	.2
<i>F+</i> %	69.2	73.9	4.7
<i>R</i>	15.1	16.7	1.6
<i>A</i> %	46.4	50.0	3.6

^a None of these differences even approaches the 5% level of significance.

from 0 to .4 and for the control group from 0 to .1. These differences proved to be without significance.

Individual Factor and Category Scores

A third method of analysis was attempted which involved the test-retest shifts of those *individual* scores which are customarily derived from each of the tests used. For example, when dealing with the Rorschach, shifts in such individual scores as *F*, *M*, number of responses, *W*%, *D*%, and *C* were studied. It was found that by and large the mean differences were small in both the experimental and control groups. None of the differences obtained even approached the 5% level of confidence (Tables 4 and 5).

Since few formal individual scores³ are customarily calculated during the process of figure drawing analysis, an attempt was made here to derive a variety of simple indices which would be comparable to the single factor Rorschach scores. The variables chosen for evaluation were size of drawings, amount of detail, view of each figure, sex of largest figure, posi-

tion on the page of the figures, sex of the figure drawn first, and age assigned to the figures. See Appendix VI for details concerning methods used in evaluating these variables. As shown in Table 6, within the control and experimental groups no instances of significant shift in scores from first to second testing were found.

The individual scores derived for the Word Association Test were as follows: (a) average reaction time of "neutral" words; (b) average reaction time on "personal" words; (c) average total reaction time; (d) average number of words that changed in content from first to second testings. Tabulations were made of degree of score shift in each of the groups. Degree of shift for all categories was small as can be seen in Table 7. As in all of the other levels of analysis of the data, the analysis of the shifts in the word association data indicated that no real changes had occurred. None of the obtained differences even approached the 5% level of confidence.

Deviants

Although the over-all results of the

TABLE 5
RORSCHACH INDIVIDUAL SCORE TEST-RETEST
MEANS AND DIFFERENCES FOR THE
CONTROL GROUP

Factors	Test	Test	Differences ^a
<i>W</i>	4.1	4.3	.2
<i>D</i>	9.6	9.7	.1
<i>Dd</i>	2.8	4.0	1.2
<i>M</i>	1.1	1.4	.3
<i>C</i>	.4	.3	.1
<i>CF</i>	1.1	.9	.2
<i>FC</i>	.7	.9	.2
<i>FY</i>	2.4	2.9	.5
<i>F+</i> %	68.5	70.1	1.6
<i>R</i>	18.7	17.9	.8
<i>A</i> %	45.5	47.1	1.6

³ The TAT does not lend itself to this type of single factor or single category analysis.

^a None of these differences even approaches the 5% level of significance.

TABLE 6
NUMBER AND PER CENT OF SHIFTS IN FIGURE DRAWING SCORES IN THE
EXPERIMENTAL AND CONTROL GROUPS

	Experimental Group (N=25)				Control Group (N=25)			
	Change		No Change ^b		Change		No Change	
	No.	%	No.	%	No.	%	No.	%
Size of Figure	9	36	16	64	4	16	21	84
Sex of Largest Figure ^a	13	52	12	48	13	52	12	48
View of Figures	7	28	18	72	5	20	20	80
Sex Drawn First	5	20	20	80	3	12	22	88
Age of Figures	3	12	22	88	4	16	21	84
Position on Page	5	20	20	80	8	32	17	68
Amount of Detail	9	36	16	64	9	36	16	64

^a Sex of largest figure was considered significant because clinically it has been observed that important aspects of an individual's dominance-submission relations with members of the opposite sex may reveal itself in the relative size of the male and female figures drawn.

^b None of the differences between the experimental and control groups was significant.

various analyses of the data indicated that neither the experimental nor the control group as a whole shifted significantly from test to retest, it seemed important to investigate those individual cases where subjects did show large changes from test to retest. In order to carry out this analysis, there was determined for each of the measures of change used in the study, the five subjects in the experimental group and the five in the

control group who showed the greatest amount of change. Then, the five subjects in each group who were found to have shown extreme changes the greatest number of times over the whole range of tests were separated out for special study. The test responses and protocols of these subjects were scrutinized in an impressionistic qualitative fashion to determine what might be common to them.

TABLE 7
MEAN TOTAL REACTION TIME, MEAN REACTION TIME TO NEUTRAL WORDS, MEAN REACTION
TIME TO PERSONAL WORDS, AND THE MEAN NUMBER OF RESPONSE WORDS WHOSE
CONTENT SHIFTS FROM THE FIRST TO SECOND TESTING IN THE EXPERIMENTAL
AND CONTROL GROUPS

AND CONTROL GROUPS				
	Total Mean Reaction Time (in seconds)	Mean Reaction Time to Neutral Words	Mean Reaction Time to Personal Words	Mean Number of Words that Shift
Experimental Group				
Test	5.6	5.6	5.7	8.5
Retest	5.0	5.1	4.9	
Difference ^a	.6	.5	.8	
Control Group				
Test	5.6	5.4	6.9	8.6
Retest	4.6	7.4	5.9	
Difference	1.0	2.0	1.0	

^a None of the differences in this table even approaches the 5% level of confidence.

It was found both in the experimental and control groups that these "deviant" subjects were, by and large, individuals who had shown a large shift in the amount of material they produced in responding to the test tasks from the first to the second testing. That is, it was found without exception that the extreme changes that they showed were a result of the fact that they responded in a very constricted fashion at the time of the first testing and then much more freely and copiously at the time of the second testing or vice versa. This change in responsiveness was generalized so that, for example, at one time there would be few responses to the Rorschach, only minimal description of the TAT cards, and a mere abbreviated sketching for the figure drawing, but at the second testing there would be much more material given all down the line for each test taken. There was no tendency for the higher degree of responsiveness to occur either at the first or second testing. This process seemed to occur in a chance fashion. It is difficult to define what lies behind such changes in degree of response. One doubts that they are due to any one factor or condition. Thus, in some instances they were a result of changes in level of cooperation. At other times they represented abrupt improvement or decline in the clinical state of a patient. In any case, there does not seem to be any over-all explanation or theory to account for those subjects who consistently showed the greatest shift from test to retest.

One may conjecture that a good segment of the deviants represent that portion of the disturbed patient population who are particularly sensitive to a psychological test situation and who express this sensitivity in terms of large varia-

bility in responsivity to the tests. These variations in responsivity would then result in large shifts in the over-all test pattern. One would be dealing with a class of instances in which the immediate emergency reactions elicited by the tests overshadowed and blotted out the more usual and long term patterns that are usually called forth by the test stimuli.

It should parenthetically be noted at this point that in none of the various analyses of the data which were undertaken were there significant differences between the variability of psychotic subjects and nonpsychotic subjects. The two groups of subjects are essentially alike in their lack of responsiveness to the situational disturbance to which they were exposed. These results argue against the possibility that the observed stability of the tests was a function of some special kind of rigidity or invariability associated with psychosis.

DISCUSSION

The results of the present study are clear cut. All of the lines of analysis used to evaluate the data have indicated that the Rorschach, TAT, Figure Drawing, and Word Association tests are stable from test to retest under the conditions that were set up here. The analysis of the data has presented some real methodological problems. There are so few methods available for expressing the results of projective tests in a fashion which is sufficiently quantitative to permit test-retest comparisons. Even more scarce are methods for quantifying projective tests which will at the same time retain some of the richness of the original data without sacrificing it to the quantification process. In the present study the problem was generally solved by using a number of modes of analysis,

varying from those based on narrow oversimplified indices to broad configurational techniques. Thus, for example, Rorschach test-retest results were evaluated at the level of the individual Rorschach factor (e.g., *F* and *M*), at the more configurational level involved in the weighted score approach, and also at the broad impressionistic level involved in the rating procedure. It is true that, although the various methods of data analysis indicated that neither of the groups shifted significantly from test to retest, there were a small number of individuals in each group who did show much variation from test to retest. However, close scrutiny of the test materials given by these subjects did not indicate any underlying consistency to their more variable behavior. Rather, the variations they showed seemed to be a function of variations in cooperativeness or clinical state that resulted in their giving a greater or smaller number of responses.

An important point to note about the results obtained in the present study is that there does not seem to be a gradient of stability among the projective tests used. None of the tests showed a significantly greater degree of instability than any of the other tests. This is interesting in terms of the fact that one finds it frequently assumed in clinical circles that tests like the Rorschach measure a relatively deeper and more stable aspect of the personality than tests like Word Associations and the Thematic Apperception Test. If there were such a differential of any proportions one would expect some evidence of it in the data here obtained.

The question arises why some previous studies have found that a significant shift can be produced in projective tests by means of special experimental condi-

tions. How is one to reconcile the negative results of the present study with the positive results of these other studies? It is not possible to give a simpler definitive answer. One can only speculate and ponder. To begin with, if one examines the results of those studies which have reported significant test-retest shifts, one is more impressed with the over-all stability rather than instability of the retest responses they obtained. That is, in almost all instances the significant shifts that have been found represent gross group trends. It is rare to find any studies which claim a degree of shift which would render a test record unrecognizable in its retest form. Now, where widespread changes in retest materials have been found (7, 9, 39, 48), they have been the result of rather unique experimental conditions. For example, it is apparently possible to produce large shifts in test responses by means of hypnotic techniques. If one takes individual subjects and works with them carefully over a period of time to the point where they are good hypnotic subjects, this unique kind of relationship can be used to produce real changes in the subjects' projective responses. Similarly, it seems possible to produce fairly widespread changes in test responses by means of powerful drugs like sodium amytal (28, 55). One would guess that it would be likewise possible to obtain shifts in test responses by subjecting the individual to the influence of any type of agent (e.g., alcohol) that seriously affects cerebral functioning. At still another level, it seems to be established that subjects can be given certain types of specific conscious instructions which will cause them to change fairly wide areas of their test responses (9, 31). For example, subjects can be influenced to change the areas of the Rorschach blots

they use or the degree to which they utilize certain content categories simply by giving them strong enough conscious incentive to do so and by giving them enough details about the workings of the test so that they know how to achieve their conscious purpose.

Yet, all of these instances in which widespread shifts in test responses have been demonstrated only serve to point up the relative stability of projective tests under ordinary conditions. Surely, it is true that one can make subjects change their responses by radically changing their physiological state or by giving them an unusual conscious set. But in the ordinary day by day use of projective tests, how often will one administer the tests to subjects who have just had a dose of sodium amytal or who are consciously aware of the fact that certain kinds of responses mean certain things and so have positive or negative value?

There is an obscurity about the basic purpose of most of the work that has been done concerning the stability of projective tests. It is often difficult to decide in terms of the experimental design he has set up just what an investigator is aiming to determine about this problem of projective test stability. There are multiple aspects to the question of projective test stability; but by and large most of the people working in the field have used the concept in a generalized way and not recognized that each is perhaps talking about a different phase of a complicated situation. Some investigators when talking about projective test stability seem to be really concerned about whether subjects can consciously fake their responses. Other investigators seem to be concerned with whether the fluctuating moods and feelings of the subject in the test situation significantly influ-

ence the responses that he gives. Still other investigators seem to be concerned with how agents that physiologically affect cerebral functioning will register their effects upon projective tests. An over-all review of the work that has been done on the problem of projective test stability suggests that the thinking underlying this work may be conceptualized around three issues:

1. There is a phase of the problem which actually focuses on the issue of whether projective tests are sufficiently *insensitive*. That is, ideally one would like projective tests to be unaffected by the temporary moods, conscious manipulations, and situational sets of subjects. Most of the studies that have dealt with the influence of special instructions, situational frustration, and the general test atmosphere upon test results have been basically concerned with determining whether projective tests are sufficiently insensitive. At this level, the stability of the projective tests refers to the fact that they do not register the effects of "extraneous" variables.

2. A second phase of the problem, one which overlaps with the first, has to do with the question to which subsystems of the organism the test shows sensitivity. Here the question is whether what is "recorded" by the test from one subsystem will remain relatively unchanged even if changes occur in the functioning of some other conceptualized subsystem. Into this category fall, for example, such studies as are concerned with the effects of drugs and food deprivation upon test responses. Is it not the primary concern here to determine to what degree changes induced in one level of organism functioning can alter certain measures of functioning at other levels?

3. A third issue revolves about the

question of whether projective tests are sufficiently sensitive to capture the important aspects of the individual's personality configuration on successive occasions. This means that the tests are expected to pick out consistently the outstanding traits of the individual despite the interfering effects of various distracting variables. If the test does not possess such sensitivity, it would not be able to present a picture of the individual that would be correlated with that individual's relatively enduring patterns.

Of course, the present study has been concerned with the issue of sensitivity. It has been the aim to determine whether projective tests are sufficiently insensitive so as not to be significantly affected by the stress growing out of temporary situational embarrassment and anxiety. The over-all results obtained certainly justify the conclusion that the Rorschach, TAT, and Figure Drawing tests are insensitive enough to extraneous and disturbing stimuli so that almost any of the conceivable special conditions arising in the course of testing the average patient or nonpatient with projective tests would not be effective in significantly distorting the data obtained. The present study can be of only limited value in clarifying the question whether "onslaught" on other subsystems of the organism, other than those here attempted, would produce greater test-retest shifts. It is possible that some dramatic acting out on the part of the examiner or stirring up of anxiety about other issues (e.g., if the patient had been told that she were going to receive shock treatment immediately after the testing) might have resulted in over-all reverberations which would have caused large shifts from test to retest. One could probably most effectively deal with this whole issue by study-

ing the test-retest patterns of subjects who were exposed to a whole range of stressful situations, differing perhaps in kind and in degree.

Finally, it would appear from the results of the present study that the projective tests are sufficiently sensitive to certain dimensions of the personality that they continue to "capture" these dimensions just about as well in the midst of a confusing background of anxiety and disturbance as they were able to under less distracting conditions.

SUMMARY AND CONCLUSIONS

1. This is a report of a study dealing with the stability of a variety of projective test productions. The primary purpose of the study was to determine whether responses to a series of projective tests would be significantly influenced by the stress growing out of temporary situational embarrassment and anxiety.

2. Fifty hospitalized female patients were studied. Twenty-five of these patients comprised a control group and 25 an experimental group. There was administered to the subjects in the experimental group a battery of four projective tests (Rorschach, TAT, Figure Drawing, and Word Association). These tests were administered immediately after the subjects had been given a disturbing gynecological examination, and they were re-administered again after five days had passed. The same battery was administered to a control group without the accompaniment of a preceding disturbing examination and was readministered again after a five day interval.

3. The data obtained were evaluated for degree of test-retest shift in terms of three different methods.

- (a) Evaluations were made by means of

impressionistic over-all ratings.

(b) Evaluations were made by a study of conventional individual factor and category scores usually derived for the given projective test.

(c) A further analysis of shifts was made in terms of a "weighted score" technique.

4. No significant changes could be found in Rorschach test responses, no matter what the level of analysis employed. Real changes could not be detected in either the control group or the experimental group.

5. Thematic Apperception Test responses also did not shift significantly from test to retest in either the experimental or control groups.

6. No significant shifts in figure drawings or figure drawing stories were observed from test to retest in the two groups.

7. A variety of word association indices and measures were found not to vary significantly from test to retest in the two groups.

8. There was no evidence among the tests used of a hierarchy of sensitivity to the temporary situational stress employed.

9. The total pattern of obtained results indicates clearly that the projective tests here studied were not significantly sensitive to the effects of the embarrassment and anxiety induced in the experimental subjects.

REFERENCES

1. ABRAMSON, L. S. The influence of set for area on Rorschach test results. *J. consult. Psychol.*, 1951, 15, 337-342.
2. ALLEN, R. M., & DORSEY, R. N. The effect of suggestion on human movement productivity in Rorschach's test. *Z. diagnost. Psychol.*, 1954, 2, 137-142.
3. ATKINSON, J. W., & MC CLELLAND, D. C. The projective expression of needs: II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, 38, 643-658.
4. BECK, S. J. *Rorschach test: I. Basic processes*. New York: Grune & Stratton, 1944.
5. BELLAK, L. *Psychology of physical illness*. New York: Grune & Stratton, 1952.
6. BERGER, D. The Rorschach as a measure of real-life stress. *J. consult. Psychol.*, 1953, 17, 355-358.
7. BERGMAN, M., GRAHAM, H., & LEAVITT, H. Rorschach exploration of consecutive hypnotic chronological age level regression. *Psychosom. Med.*, 1947, 9, 20-28.
8. BUHLER, C., BUHLER, K., & LEFEVER, D. Rorschach standardization studies: I. Development of the basic Rorschach scores with manual of directions. Los Angeles: Author, 1948.
9. CARP, A., & SHAVZIN, A. R. The susceptibility to falsification of the Rorschach psychodiagnostic technique. *J. consult. Psychol.*, 1950, 14, 230-233.
10. CHALKE, F. The Harrower Stress Tolerance Test. *Psychosom. Med.*, 1946, 8, 215-216.
11. COLEMAN, W., The TAT: I. The effect of recent experience. *J. clin. Psychol.*, 1947, 3, 257-264.
12. DE VOS, G. A quantitative approach to affective symbolism in Rorschach responses. *J. proj. Tech.*, 1952, 16, 133-150.
13. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
14. ELIZUR, A. Content analysis with regard to anxiety and hostility. *J. proj. Tech.*, 1949, 13, 247-284.
15. FELDMAN, M. J., & GRALEY, J. The effects of an experimental set to stimulate abnormality on group Rorschach performance. *J. proj. Tech.*, 1954, 18, 326-334.
16. FINE, H. J., FULKERSON, S. C., & PHILLIPS, L. Maladjustment and social attainment. *J. abnorm. soc. Psychol.*, in press.
17. FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, 64, No. 1 (Whole No. 307).
18. FISHER, S., & HINDS, E. The organization of hostility controls in various personality structures. *Genet. Psychol. Monogr.*, 1951, 44, 3-68.
19. FOSBERG, J. An experimental study of the reliability of the Rorschach psychodiagnostic technique. *Ror. res. Exch.*, 1941, 5, 72-84.
20. GUNDERSON, E., & LEHNER, G. Reliability in a projective test. *Amer. Psychologist*, 1949, 4, 387. (Abstract)
21. HARROWER, M., & GRINKER, R. The Stress Tolerance Test: Preliminary experiments with a new projective technique utilizing

- both meaningful and meaningless stimuli. *Psychosom. Med.*, 1946, 8, 3-15.
22. HERTZ, M. The reliability of the Rorschach Ink Blot Test. *J. appl. Psychol.*, 1934, 18, 512-20.
 23. HOLZBERG, J. D., & WEXLER, M. The predictability of schizophrenic performance on the Rorschach test. *J. consult. Psychol.*, 1950, 14, 395-400.
 24. HUNT, J. McV. *Personality and The Behavior Disorders*. Vol. II. New York: Ronald Press, 1944.
 25. HUTT, N. L., GIBBY, R., MILTON, E. O., & POTTHARST, K. The effect of varied experimental "sets" upon Rorschach test performance. *Ror. res. Exch.*, 1949, 13, 240-241.
 26. JOHNSON, J. C., & STERN, J. A. Rorschach rigidity and response to photic stimulation. *J. consult. Psychol.*, 1955, 19, 311-317.
 27. JUNG, C. Studies in word association. In R. S. Woodworth (Ed.), *Experimental psychology*. New York: Holt, 1938.
 28. KELLEY, D. M., LEVINE, K., PEMBERTON, W., & LILLIAN, K. K. Intravenous sodium amytal medication as an aid to the Rorschach method. *Psychiat. Quart.*, 1941, 15, 68-73.
 29. KENT, G., & ROSANOFF, A. Association. In R. S. Woodworth (Ed.), *Experimental psychology*. New York: Holt, 1938.
 30. KERR, M. Temperamental differences in twins. *Brit. J. Psychol.*, 1936, 27, 51-59.
 31. KIMBLE, G. Social influence on Rorschach records. *J. abnorm. soc. Psychol.*, 1945, 40, 89-93.
 32. KLATSKIN, E. H. An analysis of the effect of the test situation upon the Rorschach record: Formal scoring characteristics. *J. proj. Tech.*, 1952, 16, 193-199.
 33. LEVINE, K. N., GRASSI, J. R., & GERSON, M. J. Hypnotically induced mood changes in the verbal and graphic Rorschach: A case study. *Ror. res. Exch.*, 1943, 7, 130-144.
 34. LEVINE, K. N., GRASSI, J. R., & GERSON, M. J. Hypnotically induced mood changes in the verbal and graphic Rorschach: Part II. The response records. *Ror. res. Exch.*, 1944, 8, 104-124.
 35. LORD, E. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, 64, No. 10 (Whole No. 316).
 36. MCCURDY, H. Group and individual variability on the Goodenough Draw-A-Man Test. *J. educ. Psychol.*, 1947, 38, 428-436.
 37. MACHOVER, K. *Personality projection in the drawing of the human figure*. Springfield, Illinois: Thomas, 1949.
 38. MENZERATH, P. Die bedeutung der sprachlichen geläufigkeit oder der formalen sprachlichen beziehung für die reproduktion. *Zeitschrift für Psychologie*, 1908, 48, 1-95.
 39. MERCER, M., & GIBSON, R. W. Rorschach content in hypnosis: Chronological age regression. *J. clin. Psychol.*, 1950, 6, 352-359.
 40. NORMAN, R. D., LIVERANT, S., & REDLO, M. The influence of a superficial immediately preceding "set" upon responses to the Rorschach. *J. consult. Psychol.*, 1952, 11, 261-264.
 41. PIOTROWSKI, Z. A. The reliability of Rorschach's erlebnistypus. *J. abnorm. soc. Psychol.*, 1937, 32, 439-445.
 42. RABIN, A. Test constancy and variation in the mentally ill. *J. gen. Psychol.*, 1944, 31, 231.
 43. RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. I & II. Chicago: Yearbook Publishers, 1945-46.
 44. RODNICK, E. H., & KLEBANOFF, S. G. Projective reactions to induced frustration as a measure of social adjustment. *Psychol. Bull.*, 1942, 39, 489.
 45. ROMAN, R. A comparative study of northern and southern ethnocentrism as related to the rigidity of personality patterns. Unpublished doctoral dissertation, Univer. of Houston, 1953.
 46. ROSENWALD, A. A comparison of the Rorschach and Behn-Rorschach tests. Unpublished doctoral dissertation, Northwestern Univer., 1948.
 47. ROTTER, J., & RODNICK, E. A study of the reactions to experimentally induced frustration. *Psychol. Bull.*, 1940, 37, 557. (Abstract)
 48. SARBIN, T. Rorschach patterns under hypnosis. *Amer. J. Orthopsychiat.*, 1939, 9, 315-319.
 49. SLUTZ, M. The unique contributions of the Thematic Apperception Test. *Psychol. Bull.*, 1941, 38, 704.
 50. TOMKINS, S. Limits of material obtainable in the single case study by daily administration of the Thematic Apperception Test. *Psychol. Bull.*, 1942, 39, 490.
 51. TROUP, E. A comparative study by means of the Rorschach method of personality development in pairs of identical twins. *Genet. Psychol. Monogr.*, 1938, 20, 461-556.
 52. VERNON, P. Recent work on the Rorschach test. *J. ment. Sci.*, 1935, 81, 894-920.
 53. WELLS, F. L. Practice effects in free association. *Amer. J. Psychol.*, 1911, 22, 1-13.
 54. WERTHAM, F., & BLEULER, M. Inconstancy of the formal structure of the personality: Experimental study of the influence of mescaline on the Rorschach test. *Arch. Neurol. Psychiat.*, 1932, 28, 52-70.
 55. WILKINS, W. L., & ADAMS, A. J. The use of the Rorschach test under hypnosis and under sodium amytal in military psychiatry. *J. gen. Psychol.*, 1936, 131-138.

APPENDIX I

- | | |
|-------------|-------------|
| 1. Soft | *11. Nipple |
| *2. Woman | 12. Dream |
| 3. Chair | *13. Blood |
| 4. Hospital | *14. Finger |
| *5. Suck | *15. Love |
| 6. Window | 16. Plant |
| *7. Tickle | *17. Birth |
| *8. Child | 18. Curtain |
| *9. Prick | 19. Wish |
| 10. Afraid | 20. Doctor |

NOTE.—All starred words are words which are considered “personal” as opposed to the remaining 10 words considered “neutral.”

APPENDIX II

AMOUNT OF ENERGY INVESTED
IN RORSCHACH RESPONSES

The term “energy invested” refers simply to the amount of effort that the individual puts into her responses. Thus, it is a way of converting into numerical form what the subject was willing to spontaneously “offer” of herself. A number of different indices were used to measure different aspects of this “amount of effort.” These indices were based purely on clinical empirical knowledge. An example of this kind of empirical thinking is the use of number of responses to refer to the quantity aspect of effort. One assumes simply that the more responses given, the greater the amount of energy that is put out. At another level, quality and complexity of energy was judged in terms of *M* and *W*. The integration of *M* into a response or the formation of a response by integrating disparate card areas would seem to entail a degree of special organizing effort not typical of purely form responses.

The score itself is set up in terms of a

number of penalty weights for what is considered to be indicators of lack of energy invested.

1. Number of responses. A high score indicates a low degree of energy output.

10 or less responses	10
11-15 responses	7
16-20 responses	3
21-30 responses	2
31 or more responses	0

If there are fewer than 10 responses and there is not at least one “additional” response given during the inquiry 2
2. *W* per cent (for good form responses only).

0-5 per cent	10
6-10 per cent	8
11-17 per cent	5
18-25 per cent	3
26 per cent or more	0
3. *W* minus per cent (if the record has seven or more *W*'s, an added weight of two is given).

100-50 per cent	10	4 or more <i>M</i>	0
49-40 per cent	7	In each record there should be	
39-20 per cent	5	at least two responses using	
19 per cent or less	0	complex determinants (<i>FV</i> , <i>FY</i> , <i>FC</i>). Less than two	8
4. <i>W</i> plus (All <i>W</i> 's that are not <i>P</i> and are <i>F</i> [no <i>CF</i> s])		6. Perservation. (three successive responses of poor quality on the same or successive cards)	5
If none on cards 1, 4, 5	4	7. Content	
If none on cards 2, 6, 7	3	If <i>A</i> % is over 75%	3
If none on cards 3, 8	2	Any other category (except <i>H</i> and <i>A</i>) where more than three responses are in the same cate- gory	2
If none on cards 9, 10	1		
5. Complex determinants			
<i>M</i>			
Zero <i>M</i>	10		
1-2 <i>M</i>	6		
3 <i>M</i>	2		

APPENDIX III

WEIGHTED SCORE FOR EVALUATING ATTITUDE TOWARD MEN
ON THE THEMATIC APPERCEPTION TEST

Attitude toward men was scored in the following fashion:

1. In each story where the adjectives applied to men were complimentary or somehow flattering, where the men in the story were given positive, helpful, or pleasant roles, a plus (+) was given.

2. In each story where the adjectives applied to men were derogatory or where they were placed in unfavorable positions a negative (−) score was given.

3. In each story in which the attitude could not be discerned, a zero (0) score was given. Actually only those stories at the extremes which could with certainty

be judged as positive or negative were used.

The total score for this variable for each subject was the difference between the total number of positive scores and the total number of negative scores in the record.

It was felt that a great deal of subjectivity might enter into such a scoring system and consequently, to check the objectivity, all scoring was done jointly by two judges. Some instances of disagreement arose but they were settled by joint discussion.

APPENDIX IV

WEIGHTED TAT SCORE FOR OPTIMISM-PESSIMISM

This score is simply the summation of the number of stories in each record with positive endings and the number of stories with negative endings. Each of the stories was scored:

1. Plus (+) if the story ended happily and beneficially to the characters involved.

2. Minus (−) if the story ended in frustration to the main character.

3. Zero (0) was assigned to each story whose ending was unclear. The zero was also assigned to stories which did not have an ending.

The total score for optimism-pessimism was equal to the difference be-

tween the number of plus stories and the number of negative stories. All scoring was done jointly by two judges. Some instances of disagreement arose but they were settled by joint discussion.

APPENDIX V

WEIGHTED SCORE FOR EVALUATION OF AMOUNT OF ENERGY INVESTED IN THE TAT

Penalty weights were assigned to each story according to its degree of completeness. Thus, a story which included the past, present, and future, and was fairly well integrated received no penalty. At the other extreme a complete rejection of a TAT card was given a score of 7.

Weights were as follows:

Card rejection	7
Only card description	6
Assigning roles to characters but	

not making up a story	5
Story partially told. That is, a statement of the main plot, past and present, but no future or no ending, etc.	3

Complete story. Past present, future, all included	0
--	---

The final score for each subject was the summation of the penalty weights assigned each story.

APPENDIX VI

SCORING OF VARIOUS CHARACTERISTICS OF FIGURE DRAWINGS

Only those figure drawing scoring procedures are described in which the method of evaluation is not obvious from the name of the variable itself. Thus, the scoring of "sex drawn first" requires no explanation; whereas "amount of detail" does require explanation.

SIZE OF FIGURE

Change in size of figure from first to second drawing was based literally on a comparison of the size of the two figures. Very small differences were not considered significant. It was only when the difference was marked and obvious that it was recorded.

VIEW OF FIGURES

Any shifts in the view of the figures presented were recorded. That is, if the male figure was first drawn in profile and then on retest drawn to present a rear view, this was considered a shift. A change in either the male or female figure or in both figures was considered as one unit of change.

AGE OF FIGURES

Any marked changes in the age assigned to the figures by the patient in her figure drawing stories were recorded. Shifts from a child to an adult, from an

average aged adult to a very old person, etc. are examples of significant shifts. A change in either the male or female figure or in both figures was considered as one unit of change.

POSITION ON PAGE

In evaluating "position on page," the page was considered to consist of a center area and a periphery consisting of a border about the page of approximately two inches in thickness. If upon retest either of the drawings shifted from the center to the periphery or from the periphery to the center this was considered a significant change in position.

AMOUNT OF DETAIL

The following variables were considered in evaluating changes in amount of detail in figure drawings:

1. Presence of head.

2. Presence of sense organs on the head.
3. Presence of trunk.
4. Presence of arms.
5. Presence of fingers.
6. Presence of foot.
7. Presence of clothing on head (e.g., hat).
8. Presence of clothing on trunk.
9. Presence of clothing on hands (e.g., gloves).
10. Presence of clothing on foot (e.g., shoe).
11. Presence of background "scenery," furniture, or "props."

All changes involving the occurrence of any of these items in either of the first drawings and their nonoccurrence upon retest, or their nonoccurrence in the first drawings and their occurrence upon retest were considered to indicate change in amount of figure drawing detail.

Psychological Monographs: General and Applied

Expressive Graphic Movements and Their Relationship to Temperament Factors¹

MAX TALMADGE

Emma Pendleton Bradley Hospital

THIS STUDY has two main purposes: first, a theoretical purpose, that of attempting to contribute to further understanding of the unity of personality through graphic-motor movements; and secondly, to test certain hypotheses relating graphic-motor movements to an aspect of personality known as temperament.

It has long been assumed that individual characteristics of gesture and movement reveal personality differences. Often we make judgments about a person by his walk, speech, or posture. We speak about an energetic person when the walk is rapid and there appears to be purpose in his stride. When the bearing is erect, the individual seems to be sure of himself, and we speak of self-confidence or dominant attitude.

¹This study is a condensation of a dissertation submitted in partial fulfillment of the requirements for the doctoral degree at the University of Utah, Salt Lake City. The writer wishes to express his appreciation to the following members of the Departments of Psychology and Psychiatry: M. Duane Bown, for his helpful criticism and constant support throughout this investigation; William H. Brown, for his assistance in formulating the specific problem; Ija N. Korner, for his initial suggestion concerning the area of investigation; and to Don Justesen, for building the pressure recording apparatus. I owe a debt of gratitude to Anthony Davids, Chief Psychologist at the Bradley Hospital, for giving of his time and assistance in helping to organize the data, as well as the initial critical reading of the thesis, and to Elizabeth Nickerson and Christine Danat for their valuable assistance in typing the manuscript.

Movement has often been classified under the heading of expressive behavior. Allport and Vernon (4) point out that movement patterns may be broken down into expressive and adaptive parts. Allport defines expressive movements as "individual peculiarities in the manner of performing adaptive acts" (2, p. 465). Thus, it appears that, in almost every motor performance, there is individual expression.

While gross motor phenomena readily reveal differentiating characteristics of the performer, i.e., walking, bearing, and even sitting down, it has also been claimed that finer motor behavior, i.e., handwriting, drawing, and simple sensory motor performance can also reveal inner personality traits. The assumption is made that a person's movements, rather than being accidental or chance determined, are consistent under different environmental conditions, and are related to the basic motivations of the organism.

The whole area of expressive behavior is an important one in the study of personality. Within recent years, psychologists have increasingly emphasized the unity of personality. Various authors have regarded the human organism as a highly integrated, consistent being whose structure and function are meaningfully related. Moreover, they have assumed that this relationship can be determined. The psychosomatic approach to disturb-

ance of function has revealed much about the interrelationship between psycho-emotional needs and the state of the organism. The study of expressive behavior can thus be useful in understanding the structure and function of the individual organism as well as of the unity and consistency of personality.

THE PROBLEM

Historical Background

Various psychologists (11, 22, 42, 45, 59) have noted that in early infancy there exist temperamental differences. Emotional patterns in adulthood also have been found to be relatively fixed (29, 49). Johnson (27) advanced the theory that temperamental differences are indicated by different tensions. Several authors hold that muscle tension may best be studied through the hand (28, 50, 57, 58). Luria (32) devised a method for studying motor behavior under strong emotions. He found that the presence of strong emotion characteristically disorganizes the pattern of movement in the right hand and often the uniformity of tension in the left hand as well. Woodworth & Schlosberg (56) emphasize that muscular tension or tonic level is not solely a situational end product, but part of the mechanism producing alertness as well as being intimately associated with consciousness.

A variety of psychotherapists have commented upon the relationship between changes in therapy and the muscular system. Whitehorn (51) arrived at the conclusion that speech productions were overrated as indications of what is going on in a patient's mind, and used observation of motor expressive behavior in his investigations. Deutsch (14) reached the conclusion that every person had a characteristic postural pattern which represented the response of his motor apparatus to unconscious psychologic complexes. In the course of treatment, the pattern became altered. The basic nature of motor behavior is suggested by Perl: "Next to language, movement of our body or its parts is our most common means of expression. In the development of the human race, expression by movement preceded the use of organized language and the same is true of the development of the individual—it is quite possible that, because our earlier reactions were freely expressed by movement, the expression by movement continues to be more free, less inhibited than verbalization" (40, p. 598).

Expressive movements have been studied in a variety of forms. The forms of expression may

be classified under the following categories: facial expression, hands, voice, literary style, gait, painting and drawing, handwriting, and motor behavior. Bell (9) and Anderson and Anderson (6) provide surveys of studies in expressive movements. They emphasize that, although expressive movements have been the concern of psychologists for a long time, very little objective research has been carried out in this area. Maslow (33), in a recent book, makes the point that, if one wishes to know about character structure, the best behavior to study is expressive rather than coping behavior.

In studying expressive movements, three main questions have to be dealt with in order to provide a comprehensive evaluation of the investigation. First, is there consistency? Are the same results obtained with repeated testings, or are judges reliable among themselves? Secondly, is there unity to the expressive movements? There should be a meaningful relationship among the various movements to indicate that we are not dealing with isolated phenomena. Third, are these movements meaningfully related to other measurements of behavior? Is there a relationship between measurements of expressive movement and other measurements or clinical observations? A survey of the work done in this area reveals that none of the investigations meet all three criteria successfully.

Most work in the area of expressive movements is based upon impression; that is, judges make judgments about certain movements and attempt to relate them to certain personality characteristics. Eisenberg (17, 18, 19, 20), who utilized this approach and the direct measurement of expression, emphasizes that stereotyping is a large source of error in the matching techniques. Thus, the judges have a stereotype of what a "dominant" person will do and then match dominant personality characteristics to this stereotype. In other words, there is an assumption of unity in the minds of

the judges rather than a real penetration to the inner characteristics of the subjects. The importance of expression as objectively measured, rather than expression based upon impression, appears to be a vital consideration in attempting to discover whether valid relationships exist between personality variables and expressive movements.

Previous Investigations

The best known study in the area of expressive movements is that of Allport and Vernon (4) in 1933. They set out to investigate the intra-individual consistency of expressive movements by direct measurement. Using such measures of motor performance as estimation of distance with hands and feet, writing pressure, estimations of weights, intensity of voice, and number of lines on a page, they found that, for their 25 male subjects, the repeat reliabilities in the same experimental session averaged .75; those at different sessions averaged .64. They concluded that single habits of gesture were stable characteristics in their experimental group. Another question they set out to answer was whether different movements from different parts of the body, performing different functions, correlate highly. On the basis of their findings, they concluded that there was no general overall movement factor. An analysis of their intercorrelations led them to the conclusion that there were three group factors, "areal," "centrifugal," and "emphasis." This classical study gives the first experimental indication that bodily movements are "meaningful"; that is, that expressions of the body may be used to better understand the nature of the individual personality. However, Allport and Vernon left off at the most crucial point. They failed to study the relationship between bodily movements and inner personality structure.

Using graduate students, Eisenberg (18) repeated Allport and Vernon's work and obtained similar results. Enke (21) used 500 subjects and found characteristics differentiating the movements of pyknic and nonpyknic types. Arnheim (7) and Wolff (52, 53, 54), in a large number of studies, matched different types of expressive movements through subjective impression. On discovering consistency in impression, they concluded, by inference, that there is a corresponding consistency in expression.

A number of investigators have dealt with graphic-motor expression. Many books have been written about graphology and drawings as indicative of personality and individual differences.

The lack of objective studies and the difficulty in determining specific measurable categories in both these areas have limited the validity of these techniques in experimental investigations. As far back as 1924, it was claimed by Poffenberger and Barrow that, "The moment we touch upon line form we are already, in strictness, beyond the elements. For with form enters the motor factor, which cannot be separated from the motor innervations of the whole body. . . . The impression of the lines involves expression, a meaning which we cannot escape" (41, p. 102).

Various experimenters (25, 31, 41) have found a certain uniformity in depicting emotional states by direction and type of line drawn. Lundholm (31), for example, found that subjects drew lines symbolizing states of strong motor expression by short waves and acute angles, and lines symbolizing states of weak motor expression by long and low waves. He also found that the downward tendency of a line expresses relaxation; the upward tendency, power. Poffenberger and Barrow (41) had their subjects match lines with 47 adjectives. They found that slow, descending curves were related to sad, lazy, and weak; slow horizontal curves, to quiet and gentle. Harms (25) utilized a much larger population and found that subjects gave similar meanings to identical form patterns. Some of his findings were: depressed = line downward; laughing = line upward; fear = downward, bent lines; tired = horizontal line going downward; sad = horizontal and vertical lines downward.

Downey (15, 16) was one of the first to use objective measures of handwriting as reflecting personality traits. In one study, she obtained high positive correlations between the trait, "attention to detail," and such measures as smallness of letter, regularity, and invariability of spacing. With respect to the trait, "pride" or "feeling of self-worth," she found high correlations with size of letter and ornamentation of capitals. With the trait rating continuum "explosive-inhibited" the following group of handwriting characteristics were highly related—exaggerated size, rapidity, and pressure.

In a more recent study, Pascal (39) attempted to find quantifiable variables of handwriting and to relate them to five personality traits as rated by seven psychologists. He found that the correlation between pressure range and energy was significant at the .01 level of confidence. Speed of writing and pressure range was significantly correlated with impulsiveness (.35), pressure range was significantly correlated with energy and expression (.62), and speed was significantly correlated with expression and impulsiveness (.54).

Various investigators have claimed that motor behavior reveals emotional status. Munsterberg

(38), Allport (1), and Downey (16) found that there are considerable individual differences in the subjective estimation of space according to changes of mood and purpose. Wolff (55) related that Remmers and Thompson asked students to draw lines on paper while thinking of pleasant events and unpleasant events. The lines drawn during pleasant thoughts were significantly longer than those during unpleasant thoughts. The involuntary tendency toward extension or contraction of movement appeared to be based upon emotional stimuli.

Based upon some of the investigations mentioned previously, Mira (34) constructed a principle of psychomyokinesis which holds that "Psychological space is not neutral; all the movements of the individual acquire a peculiar meaning according to the direction in which they are performed" (p. 175).

Mira further postulates that mental activity may in turn be considered a succession of acts implying postural changes and that, whenever mental equilibrium is disturbed, there is a corresponding distortion of movements, dependent upon the degree to which any voluntary attempt at compensation is eliminated. He consequently claims that, if an individual is requested to perform small oscillatory movements in the fundamental direction of space without visual control, the observed shifts will indicate the predominant muscular pattern and will furnish insight as to the subject's fundamental attitudes of reaction. In addition, he assumes that the differences so detected may signify the individual's habitual attitudes of reaction provided they are sufficiently constant.

In line with the above, Mira devised a test called the Myokinetic Psychodiagnosis technique (MPD). Briefly, this requires subjects to trace lines and other figures while their vision is screened. The lines to be drawn are known as "lineograms" or "kineotograms." They are drawn with the right and left hands in three different planes: horizontal, vertical, and sagittal (perpendicular) to the front of the body. Other figures, including zigzag lines, staircases, circles, chains, and parallels are also drawn, supposedly to check the initial lineograms. Objective scores are derived from length of lines, direction of shiftings, comparison of movements towards and away from the body, and other relationships. It should be pointed out that, unlike the situation in handwriting or drawing, the subject is required to be screened and to keep his hand and arm always in the air, never in a resting position.

Investigations using this procedure are very few and even the studies that Mira has done, rather than using objective data, consist mainly of conclusions based upon poorly designed observations (5, 35, 36, 37). It is evident that any worthwhile study using the Mira technique

will require a set of hypotheses that can be tested. Moreover, reliability measures will need to be indicated. Bell writes of this test that, "In its present state of development, its results are so tentative that it cannot be applied immediately for discrimination between clinical groups or for analysis of the structure of the personality in individual cases. Standardized experiments with the technique should be relatively easy to develop, since the major analysis of the results of its use by various experimenters should be possible" (9, p. 340).

The Present Investigation

The present investigation is designed to determine whether graphic-motor movements are reliable and internally consistent, and to test whether these movements are meaningfully related to temperamental characteristics. A modification of the Mira technique is used as the test instrument. Mira's test has not been validated and his claims have not been substantiated; hence, little use is made of previous findings. In the present modified form of this test, only some of the quantitative scorable categories suggested by Mira are used. Other measures not used by him are also utilized.

Personality is an all-encompassing term, dealing with every aspect of human behavior. A review of the literature suggests that more meaningful relationships could be discovered by limiting the area of investigation to that aspect of personality known as temperament. Temperament has usually been broken down into three aspects: emotional, psychological, and kinetic. A definition giving due weight to all three functions of temperament is the one by Allport:

"Temperament refers to the characteristic phenomena of an individual's emotional nature, including his susceptibility to emotional stimulation, his customary strength and speed of response, the quality of his prevailing mood, and all the peculiarities of fluctuation and intensity in mood; these phenomena being regarded as dependent upon constitutional make-up, and therefore largely hereditary in origin" (2, p. 54).

Some writers (10, 16) have emphasized the kinetic or motor discharge aspect of temperament. This concept is especially pertinent and suggests that graphic-motor movements may profitably be compared with questionnaire data on temperament.

After surveying the available objective instruments for assessing temperament, it was decided to use the Thurstone Temperament Schedule (46, 47). This is a test based upon a second order factor analysis of Guilford's thirteen personality factors (23, 24). The theory is advanced by Thurstone that second-order factors are more basic and pervasive characteristics of the original behavior items included in Guilford's original factorial study. It is postulated that each of the factors is a continuous variable and that each occurs in every member of the population but in varying degrees of strength.

The test appears to have been well standardized and retest reliabilities vary from .61 to .82, depending upon the particular factor. The following factors were defined:

Active (A). A person scoring high in this area usually works and moves rapidly. He is restless whenever he has to be quiet. He likes to be "on the go" and tends to hurry. He usually speaks, walks, writes, drives, and works rapidly, even when these activities do not demand speed.

Vigorous (V). A person with a high score in this area participates in physical sports, work requiring the use of his hands and the use of tools, and outdoor occupations. The area emphasizes physical activity using large muscle groups and great expenditure of energy.

Impulsive (I). High scores in this category indicate a happy-go-lucky, daredevil, carefree, acting-on-the-spur-of-the-moment disposition. The person makes decisions quickly, enjoys competition, and changes easily from one task to another.

Dominant (D). People scoring high on this factor think of themselves as leaders, capable of taking initiative and responsibility. They enjoy public speaking, organizing social activities, promoting new projects, and persuading others.

Stable (E). Persons who have high Stable scores usually are cheerful and have an even

disposition. They can relax in a noisy room, and they remain calm in a crisis. They claim that they can disregard distractions while studying.

Sociable (S). Persons with high scores in this area enjoy the company of others, make friends easily, and are sympathetic, cooperative, and agreeable in their relations with people.

Reflective (R). High scores in this area indicate that a person likes meditative and reflective thinking and enjoys dealing with theoretical rather than practical problems. Self-examination is characteristic of reflective persons. These people are usually quiet, work alone, and enjoy work that requires accuracy and fine detail.

1. *Formulation of Relationships between Graphic-Motor Movement and Temperament Factors.* Utilizing the temperament factors as criteria, the following hypotheses have been formulated between these factors and graphic-motor variables.

(a) *Active.* From Thurstone's description, it is rather evident that time scores should provide a measure of "active" temperament. However, as we are attempting to get at natural tempo and not set up a situation where speed is stressed, time should be taken without the subject's knowledge and speed should not be mentioned. Some investigators have claimed that a relationship exists between speed and pressure (43, 54), and therefore it may be expected that pressure will also be related to the "Active" factor.

(b) *Vigorous.* The concept of energy level in Thurstone's description readily suggests that pressure may be a measure of the factor. Downey (16) equates pressure with psychic tension and such concepts as "will" and "energy." Wolff (55) related handwriting pressure to vigor, force, and vitality. Jacoby (26) found that depressive children have low writing pressure. Allport and Vernon (4) found that measures of pressure, intensity, forcefulness, and fewness of parallel lines were highly loaded in the emphasis factor. It therefore seems that fewness of

parallel lines may also be a good measure of this factor. As has been mentioned, time scores have been claimed to be related to pressure and will also be utilized. Thurstone (48) found in his apparent movement test that people who score high on the "vigorous" factor also saw movement in the leftward direction.

(c) *Impulsive*. Thurstone indicates that this factor is a measure of acting-on-the-spur-of-the-moment. Therefore, in graphic-motor movement, when shifts have to be made, individuals who score high on this factor will shift sooner than others. Thus, we can expect short lines when many lines are required. An individual who scores high on this factor might be called a nonperseverative. He does the task as fast as possible and is concerned no more with it. In addition to fast shifting, which may be reflected in short lines, speed and pressure also seem related. Pascal (39) found significant relationships between "impulsive" trait and pressure.

(d) *Dominant*. We usually think of dominant people as assertive, outgoing, and expansive. Therefore, we might expect that people who score high on this factor will shift upward in the line drawings, have larger movements in getting to a certain point, and will exert more pressure in their movements. Eisenberg (17) found that dominant students exerted greater pressure, worked faster, and were less persistent.

(e) *Stable*. It is rather difficult to relate this concept to graphic-motor movements. Thurstone found that stable people perceived movement in an upward direction. It might be speculated that emotionally stable people will have more upward movements, work faster [Roman (43) believes that a happy mood releases and accelerates the flow of movement,

whereas an anxious mood or inner tension restrains it], and will make more outgoing movements.

(f) *Sociable*. This concept appears to be related to extraversion. Various graphologists and drawing analysts claim that a predominance of rightward movement indicates a person predominantly responsive to stimulations activated from promptings from without. Thurstone (47) found that persons who perceive motion in an upward direction were socially outgoing. It is, therefore, postulated that high scorers on this factor will have upward and rightward shifting.

(g) *Reflective*. In many respects this is the opposite of impulsive and sociable. Thus, reflective individuals have been described as being perseveraters (30) and introverted. It is, therefore, postulated that longer time scores, a predominance of leftward movement, and slower shifting will be related to high reflection.

Hypotheses to be Tested

1. Graphic-motor movements are stable measures, and test-retest correlations will reveal a high degree of reliability.

2. Individuals are internally consistent in their graphic-motor movements. Similar movements will be found highly correlated.

3. Certain types of graphic-motor movements go together and orderly relationships exist among the various movements.

4. Specific relationships between the graphic-motor movements and temperament factors are postulated:

- (a) High scorers on the Active factor will work fast and exert greater pressure.

- (b) High scorers on the Vigorous factor will be found to have greater pressure, fewer parallel lines, and faster time scores.

(c) High scorers on the impulsive factor will shift sooner, work faster, and have greater pressure.

(d) High scorers on the Dominant factor will shift upward, have larger movements in reaching a certain distance and will exert more pressure.

(e) High scorers on the Stable factor will shift in the upward direction, work faster, and will make more outgoing movements.

(f) High scorers on the Sociable factor will shift in the rightward and upward direction and have outward movements.

(g) High scorers on the Reflective factor will work slower, shift slower, and will have leftward or inward movements.

METHOD AND PROCEDURE

The Subjects

The experimental data for this study were collected from the results of the modified Mira technique and the Thurstone Temperament Schedule, which were administered to 80 male college students taking psychology courses at the University of Utah. They were informed that participation in the experiment had no bearing on the class grades and that the information obtained would be used solely for research purposes and would not in any manner be made public to their instructor, to the department, or to the University.

In order to obtain as large a sample as possible and not be concerned with sex differences, only male students were selected. The variable of handedness was controlled, since only right-handed subjects were used. The 80 subjects ranged in age from 18 to 35 years, with a mean age of 21.6 years.

Each subject was seen individually and was told the following: "I am conducting a research study in which I have subjects draw lines and other designs. I am not interested in you as

individuals but only in group results. It is important that you do the best you can or else the research cannot succeed."

The Test Instrument

The test instrument is a modified version of the Mira technique. The test stimuli comprise six sheets of paper on which various designs are printed. See Fig. 1 for a reproduction of the stimulus figures. The first sheet contains a staircase with vertical and horizontal lines 5 mm. long. The second sheet contains a circle 25 mm. in diameter. On the third sheet there are three lines—a horizontal line 38 mm. long, and two vertical lines 42 mm. long. The fourth sheet has four zigzag lines, two on the bottom of the page 17 mm. apart, and two on the top of the page 122 mm. apart. The fifth sheet has two parallel lines, each 146 mm. long and connected at the bottom with two lines each 23 mm. long and 9 mm. apart. The sixth sheet is like the fifth, except that the parallel lines are connected

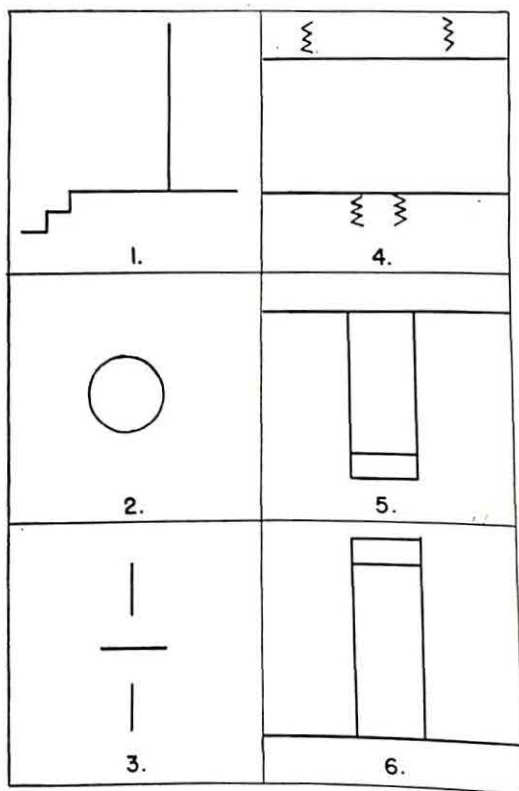


FIG. 1. Myokinetic psychodiagnosis stimulus figures. These are reproductions of the stimulus figures as they are presented to the subjects: 1. Ascending and descending staircase; 2. Circle; 3. Lines: sagittal, horizontal, and vertical; 4. Up and down zigzags; 5. Upgoing ladder; 6. Downgoing ladder.

at the top of the sheet rather than at the bottom.

Each sheet of paper was separately inserted and centered in a wooden board 15 inches long and 12 inches wide. Attached to the board, about 10 inches above it, was another board six inches long and eight inches wide. This board was used as a screen and could be moved into different positions. The larger board was affixed to a table the top of which was 30 inches above the floor. The board was slightly inclined to the left and its front was even with the edge of the table. The board could be set either in a horizontal position (down), or in a sagittal position parallel to the subject (up). Each subject was seated on a stool which could be either raised or lowered. The subjects adjusted the stool according to the most comfortable position for performing the various graphic movements. The stool was set at a fixed distance of six inches from the edge of the table.

Once the subject was comfortably seated, he was given the following instructions: "I want you to draw some lines. This will be somewhat different from your usual writing, since you may not rest your hand or elbow anywhere. Hold your arms up from your sides. Let them move freely." The subject was then given the first stimulus sheet containing the staircase. The board was in the sagittal position for this task. There were three model staircases which the subject was instructed to trace with his vision exposed. Then the subject was instructed to continue making staircases going up and to the right, until told to stop. When the subject was at the center of the page, he was instructed to stop, lift his pen, and then to come down to the right. He was told to try and make all the steps equal when going up and coming down. His vision was screened after he completed the model staircases.

The first stimulus sheet was taken away and the second one, the circle, was presented. The board remained in the sagittal position. The instructions for this task were similar to the staircase task. Each subject was required to draw ten circles with his vision screened after he traced one circle while looking at the model. Again he was reminded to keep his arms and hands off the paper and to try to keep the pen perpendicular to the paper. He was requested to draw ten circles the same size as the model and one on top of the other. Each time he completed a circle, he was to lift his hand off the paper and then proceed with the next circle.

The third stimulus sheet contained the three lines. The instructions were similar to the other tasks. The subject was told, "Here is the first line (the examiner points to the bottom vertical line). Take the pen in your hand and retrace this line going from top to bottom and then

back up again while looking at the line. Then I am going to screen you, and you will trace this line ten more times, five times alternately each direction. Try to draw all the lines the same length as the model and in the same place, one on top of the other. Remember to keep your arms and hands off the board."

Each subject was permitted to trace one line back and forth with his vision unscreened and then ten lines, five back and five forth with his vision screened so that he could not see what he was drawing. As he could not rest his hands on the paper, the pen point was always fairly perpendicular to the paper.

When the subject completed the vertical line, he was requested to proceed in the same fashion for the horizontal line, reproducing five right and five left lines. Before starting, he adjusted the stool to the most comfortable position for him. After the subject completed the ten lines in the horizontal direction, the board was raised so that it was parallel to the subject and he was instructed to make five up and down movements on the top vertical line (sagittal plane). Thus, in all, each subject made 30 lines with his vision screened.

The fourth sheet contained the zigzags. The board was put down on the table and the subject was given two pens, one in each hand. The subject was first instructed to trace the bottom zigzags (17 mm. apart). These instructions were: "Retrace these sample lines, moving both pens at the same time, inward and outward, just as if you were playing an accordion. When you have finished the samples, I will place the screen before your eyes. You will continue to draw the zigzag lines, trying to keep them the same size as the sample and in the same direction. Remember to hold your wrists up so that your hand and arm do not touch the paper. You will be told when to stop."

When he completed the bottom zigzags, he was instructed to do the top zigzags. The instructions were the same for this task, with the exception that he was told to start at the top of the sheet and work his way down until told to stop.

The fifth sheet contained the egocifugal ladder (away from the body). The writing surface remained in the horizontal position (perpendicular to the subject). Each subject was given the pen and directed to trace the two model lines beginning at the left and moving to the right. After he completed the first model line, he was instructed to lift the pen and go back to the left and trace the second line going from left to right. Then his vision was screened and he was instructed to keep making lines, trying to keep them the same distance apart as the two model lines. He was further instructed to make lines from left to right. After completing a line, he

was told to lift his pen and start at the left for the next line. He was told that it was like a ladder and he should try to keep the lines within the sides of the ladder. He was to keep making lines until instructed to stop.

The sixth sheet contained the egocipetal ladder. The instructions for this task were the same as for the fifth sheet. The model lines were at the top of the page and the subjects were directed to draw lines from the top toward himself. Again the lines were drawn from the left to the right.

The entire Mira Procedure took about twenty minutes. Figure 2 contains a reproduction of a test record.

Data Obtained During Investigation

To the right of the subject, a screen was placed so that he was unable to see the recording apparatus. During the experiment, time and pressure were recorded for each task. The time was recorded in seconds by means of a stop watch and was taken from the moment he completed the model figures until he was instructed to stop. Thus, in all there were ten different time scores, one for each of the ten tasks.

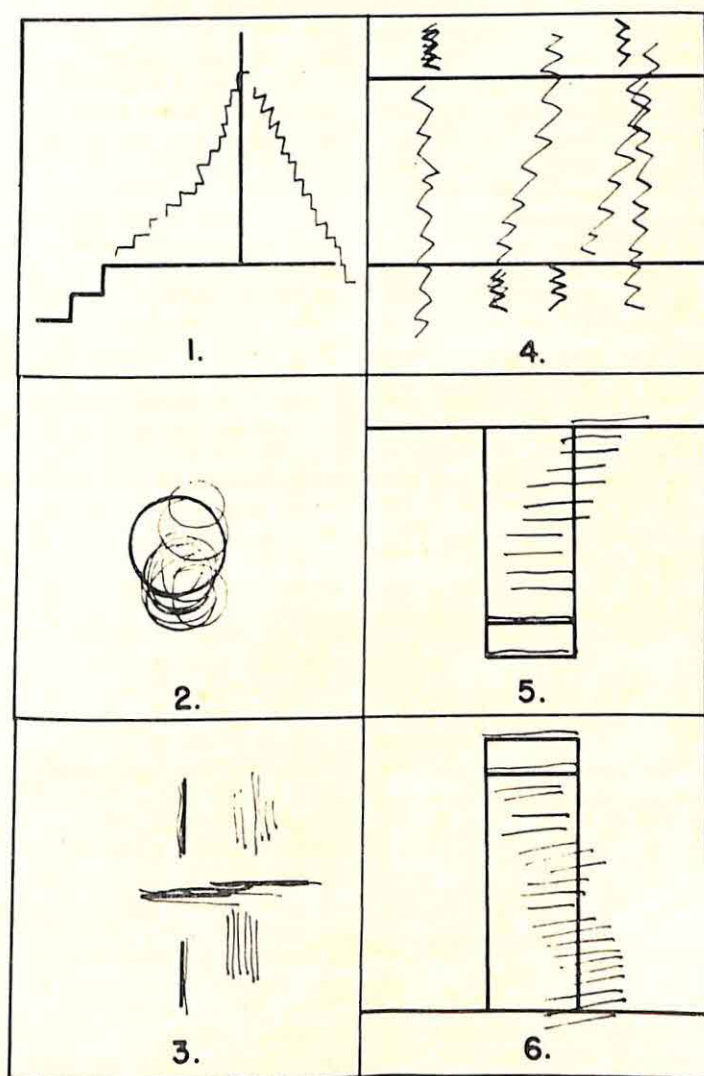


FIG. 2. Myokinetic psychodiagnosis test completed productions. These are reproductions of the drawings made by a subject.

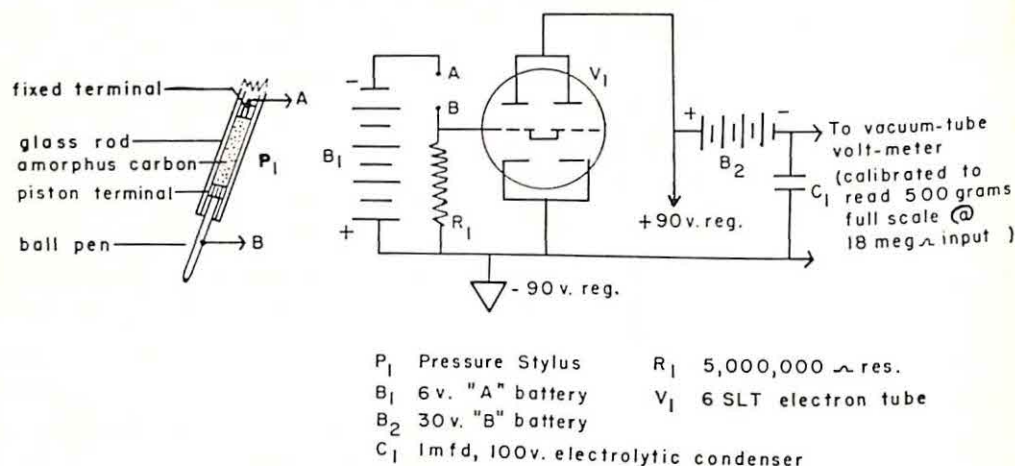


FIG. 3. Schematic diagram of pressure measuring apparatus.

Pressure was recorded by means of an apparatus devised by Don Justesen, Laboratory Assistant in the Department of Psychology, University of Utah. Figure 3 is a schematic diagram of this apparatus.

The ball-point pen that the subjects used is an electronic stylus connected to a direct-coupled amplifier and a vacuum tube volt-meter. The electronic stylus is composed of a piston which engages a pocket of carbon in a glass rod. Carbon has the unique quality of changing resistance when pressure is applied. This change of resistance causes a voltage differential upon the grid of an electron tube. Thus, changes in pressure cause changes in resistance which, in turn, produce voltage changes which are amplified by a D.C. amplifier and then fed into the vacuum tube volt-meter which has been calibrated to a full scale of 500 grams.

The use of the amplifier, in addition to the vacuum tube volt-meter, is necessary because the total swing in potential produced by the pressure stylus is less than 30 millivolts. Calibration of the scale was derived by first applying the stylus on a gram weight scale. When the scale registers 500 grams, the vacuum tube volt-meter is calibrated to full scale deflection. The position of the vacuum tube volt-meter indicator is then noted and checked for position. This is repeated with different weights until a linear scale (on the vacuum tube volt-meter) has had its graduations identified with a succession of weights.

After the experiment was completed, the recordings of pressure which were obtained from the volt-meter were transformed into grams. Thus, electronic measurements were changed into weight measurements. A full scale deflec-

tion of 5 units was equal to 500 grams, 4 units to 420 grams, 3 units to 300 grams, and so on until the scale was transformed into gram measurements. This apparatus has the unique distinction of permitting a standard piece of equipment, a ball point pen, to be applied in any direction and the obtaining of quickly interpretable pressure recordings. In addition, the volt-meter averages the pressure so that what is recorded is the average pressure for a given movement. Thus it was unnecessary to take constant recordings for each task; one was sufficient to obtain an average pressure measurement for any particular task.

Utilizing this apparatus, ten different pressure scores were obtained from each subject, one for each task.

Data Obtained from Test Instrument

From the lineogram recordings, the following data were obtained: length of line, primary shifting (up and down on the vertical lines, right and left on the horizontal), secondary shifting (right and left on the verticals, up and down on the horizontal). Calipers and a metric ruler were used to measure the length of lines and the shiftings. The length of the ten lines for each task was measured and averaged. This provided the average length for each task for each subject. Primary shiftings were obtained on the verticals by using the midpoint of the model line and measuring the length of each line above and below the midpoint. If the length was greater above the midpoint, it was considered a positive shifting, and the difference in mm. was recorded. If the length was greater below the midpoint, it was considered as negative shifting, and the difference between the length above and below the midpoint was recorded as the amount of negative shifting. For

the horizontal line, the same procedure was followed except that the measurements were taken in reference to right or left of the midpoint. Where the length was longer to the right, it was considered as positive shifting; to the left it was considered as negative shifting.

Secondary shifting was measured on the vertical lines according to the distance the lines were from the line. If all the lines were to the right of the model line, the distance was averaged and a positive score was obtained; if they were all to the left, an average negative score was obtained. There were a few in which shiftings were to the right and left of the given line. In this case the average distance in one direction was subtracted from the average distance in the other direction; the side of the higher average obtained the sign, whether positive or negative. On the horizontal line, secondary shifting was measured in terms of distance above or below the given line. The ten lines were averaged; the average positive or negative shiftings were recorded. All measurements were taken from the midpoint of the model line.

On the fourth task, the circles, similar measurements were recorded: size, primary and secondary shifting. The diameters of the ten circles were measured and averaged to give the mean circle size for each subject. Primary shiftings were recorded from the average right or left shiftings of the circles. At the midpoint of the circle, perpendicular lines were erected, and the distances from the midpoint of the given circle to the center of the other ten circles were measured. The right or left deviations were measured and averaged. The deviations were considered positive if to the right and negative if to the left. Secondary shifting was recorded in a similar manner with the measuring of up and down shifting; those that averaged above the given midpoint were considered as positive and those below negative.

There were two staircase tasks, ascending and descending. From the end of the model staircase to the left side of the paper was 17 mm., the midpoint was 85 mm. At the midline, a perpendicular line was erected. When the subject reached this line, he was told to stop and to start coming down. The ascending staircase comprises the upward horizontal and vertical lines, while the descending staircase consists of the horizontal and vertical lines to the right of the midline to the edge of the page. The measurements obtained from both tasks were similar: number of stairs (to the midline in the ascending, and back down to the level of the model line in the descending), and length of vertical and horizontal lines. The length of the horizontal and vertical lines for each task was averaged separately, and the average vertical and horizontal lengths for each subject on both tasks were obtained.

For the zigzag tasks, three measurements were recorded. The first measurement was the number of angles it took the subject to get from the model zigzag to the other end of the paper where the other model zigzag ended. This was a distance of 205 mm., in either direction going up or down. The second measurement was width, or distance apart, between the zigzags made by two different hands. This width was measured three times, one-third the length, two-thirds the length, and at the end of the distance, 205 mm. up or down, depending upon whether the up-going or down-going zigzags were being measured. The third series of measurements were deviation scores, the distance the right-handed zigzag deviated from a line drawn straight up from the end of the model zigzags. The deviations were measured at one-third intervals and three measurements were made of each line. If the average deviation were to the right of the drawn line, a positive deviation was recorded and the average deviation of the three recordings was the score obtained. The identical procedure was followed for both the up-going and down-going zigzags.

The ladder tasks provided four different types of scores. The first was the number of lines it took the subject to reach the top or bottom, depending upon whether the egocifugal ladder (away from the body) or the egocipetal ladder (toward the subject) was being measured. The second measurement was the distance between the lines. The two model lines were 9 mm. apart, and the subjects were instructed to attempt to continue making the lines the same distance apart as the model lines. The distance between the lines were measured at the midline of the ladder, averaged and divided by the number of lines to give the average distance between the lines. The third measurement was the length of lines. All the lines were measured and the average length of line was obtained by dividing the number of lines by the length of all the lines. The fourth score obtained was a deviation score. From the midline of the ladder (vertical line), the length of each line to the right and left of the midline was measured. The shorter part was subtracted from the longer and a positive or negative value was assigned, depending upon whether the right or left side was longer. All the deviation scores were averaged and an average deviation score was obtained. The same procedure was followed for the egocifugal and egocipetal ladders.

To summarize briefly: in addition to the measurements noted, there were also ten pressure and ten time scores, one for each task: vertical line, horizontal line, sagittal line, circle, ascending staircase, descending staircase, zigzag upward, zigzag downward, egocifugal ladder, and egocipetal ladder. Adding up all the measure-

ments, 52 different scores were obtained from the graphic-motor movements.

Criterion Test—Thurstone Temperament Schedule

After the subject had completed the Mira test, he was taken to another table and was requested to answer the "yes" or "no" questions on the Thurstone Temperament Schedule.

The Thurstone test contains 140 questions, which can be answered with a cross under "yes," "no," or "?." The subjects were requested not to use the question mark category if they could possibly help it. They were again reassured that we were not interested in them as individuals but only as subjects in an experiment, and the results of their test would not be made known to anyone; it was therefore very important that the questions be answered as truthfully as possible.

The Thurstone Temperament Schedule (TTS) was devised for the purpose of describing how normal people differ from each other on stable traits of personality. Thurstone emphasized that the scale does not attempt to appraise the degree of conflict, insecurity, or maladjustment. It is designed to assess those traits which are relatively permanent for each person. Thurstone acknowledges that, while the schedule is limited in depth, it has unusually broad coverage.

The test itself took about 15 to 25 minutes to complete. Scoring was very simple; all that was necessary was to count up the correct answers for each area. These are automatically recorded on the grid page, as the subject marks his answers. Raw scores were used rather than percentages, as we were not interested in standards but only relative position, that is, who performed higher and who, lower.

After the subjects had completed the Mira technique and the Thurstone Temperament Schedule, they were requested to provide additional information such as age, height, and weight and were requested to make an appointment for more testing. Out of the 80 subjects, 74 were retested on the Mira techniques over a two-week period. Test-retest reliabilities on the 52 graphic-motor variables were obtained from the test instrument.

RESULTS

As in any investigation where relationships are of major interest, it is first necessary to determine whether the variables involved are consistent. Can the same results be expected with repeated testings? This is especially true for an area in which little previous investiga-

tion has been carried out, such as in that of graphic-motor movements. It was unnecessary to test for the reliability of the Thurstone measures, as Thurstone provides the reliabilities for his scale for a similar group.

Hypothesis 1: Graphic-motor movements are stable measures, and test-retest correlations will reveal a high degree of reliability.

As previously mentioned, 74 of the original 80 subjects returned for retesting on the Mira technique. Test-retest Pearson r 's were performed on the 52 graphic-motor variables for the 74 subjects. Table 1 presents these data and reveals some degree of consistency. There was only one coefficient below .30; many of the correlations were in the seventies and eighties. Although many of the highest correlations were in the obvious physical dimensions of time and pressure, there was also a number of high correlations in such variables as: number of angles, number of lines, and length. While, in general, the pressure and time variables had the highest reliability, it was also observed that, in one task, the descending staircase, the reliabilities were greatly reduced. To obtain the average reliability for all the 52 variables, the reliability coefficients were converted into "z" scores and averaged. This gave a "z" of .720 or an "r" of .617. None of the tasks took more than a few minutes. It can be expected that, with longer tasks, the reliability would be even higher (48). These results indicate that graphic-motor performances are fairly stable characteristics for the present subjects.

Hypothesis 2: Individuals are internally consistent in their graphic-motor movements. Similar movements will be highly correlated.

All the variables were intercorrelated

TABLE 1

TEST-RETEST RELIABILITY COEFFICIENTS ON
THE 52 GRAPHIC-MOTOR VARIABLES
FOR 74 SUBJECTS

Task	Variable Number and Name	Reliability Coefficient
Vertical Line	1. Pressure	.729
	2. Time	.670
	3. Length	.717
	4. Primary Deviation	.441
	5. Secondary Deviation	.470
Horizontal Line	6. Pressure	.753
	7. Time	.731
	8. Length	.592
	9. Primary Deviation	.344
	10. Secondary Deviation	.261
Sagittal Line	11. Pressure	.766
	12. Time	.708
	13. Length	.455
	14. Primary Deviation	.387
	15. Secondary Deviation	.470
Circles	16. Pressure	.719
	17. Time	.823
	18. Length	.585
	19. Primary Deviation	.415
	20. Secondary Deviation	.412
Ascending Staircase	21. Pressure	.819
	22. Time	.526
	23. Number of Stairs	.640
	24. Vertical Length	.447
	25. Horizontal Length	.497
Descending Staircase	26. Pressure	.391
	27. Time	.408
	28. Number of Stairs	.642
	29. Vertical Length	.632
	30. Horizontal Length	.546
Uppgoing Zigzags	31. Pressure	.708
	32. Time	.813
	33. Number of Angles	.739
	34. Distance Apart	.425
	35. Deviation	.465
Downgoing Zigzags	36. Pressure	.700
	37. Time	.753
	38. Number of Angles	.775
	39. Distance Apart	.655
	40. Deviation	.514
Uppgoing Ladders	41. Pressure	.693
	42. Time	.772
	43. Number of Lines	.727
	44. Distance between Lines	.649
	45. Length	.549
Downgoing Ladders	46. Deviation	.380
	47. Pressure	.819
	48. Time	.779
	49. Number of Lines	.640
	50. Distance between Lines	.616
	51. Length	.413
	52. Deviation	.340
Task	Variable Number and Name	Reliability Coefficient

by means of I.B.M. In all, there were 1,891 separate Pearson's "r" intercorrelations; this comprises 52 graphic-motor variables obtained at the first testing, 7 temperament scores, and the variables of age, height, and weight. Table A² contains this 62 variable intercorrelation matrix. In order to pick out the significant relationships, the .05 level of significance was selected for the two-tailed test. It was found that for an N of 80, all correlations above .219 would be significant at this level. From the intercorrelation matrix, 425 out of the 1,891 correlations were found to be related at .22 or above, regardless of sign.

The problem immediately raised is: could such a large number of significant relationships be expected on the basis of chance alone? Statistically there is no way of determining what effect chance has on a large intercorrelation matrix. At the .05 level of significance, 5 out of 100 can be expected to be significant by chance alone. For 1,900 correlations, 90 significant correlations can be expected due to chance. Almost five times that number were obtained. It seems highly unlikely that the large number of significant relationships obtained is determined to any great extent by chance.

So far, all that has been indicated is that there is a "meaningfulness" to the domain covered by the intercorrelation matrix. It is evident that the area encompassed is not a helter-skelter one in which the variables are completely independent of each other. As it has been found that the number of significant relationships is significantly greater than would be expected by chance, it is now possible to examine the significant cor-

²The table of intercorrelations of the 62 variables used in the study has been deposited with the American Documentation Institute. Order Document No. 5754, remitting \$1.25 for 35-mm. microfilm or \$1.25 for 6 by 8 in. photocopies.

TABLE 2
INTERCORRELATION OF THE PROFILE SCORES REPORTED BY THURSTONE (T)
COMPARED WITH THE PRESENT STUDY (P)

	Active		Vigorous		Impulsive		Dominant		Stable		Sociable	
	T	P	T	P	T	P	T	P	T	P	T	P
Active												
Vigorous	.23*	.17										
Impulsive	.36	.47**	.36**	.47**								
Dominant	.10	.39*	.24*	.16	.37**	.49**						
Stable	-.03	.12	.09	.15	.21	.30**	.20	.30**				
Sociable	.22*	.26*	.20	.24*	.55**	.51**	.57**	.72**	.17	.37**		
Reflective	-.03	-.24*	-.02	-.06	.09	-.24*	.08	-.03	.07	-.16	-.14	-.17

* Significant beyond the .05 level.

** Significant beyond the .01 level.

relations in order to determine what relationships exist.

Before going into a detailed analysis of the graphic-motor variables, it may be best to first examine the intercorrelation of the temperament factors. Table 2 contains a comparison of the intercorrelations presented by Thurstone and those of the present study. As can be seen in most cases, higher correlations were obtained in the present investigation than were obtained by Thurstone. Such correlations as .72 between dominant and sociable, .51 between sociable and impulsive, and .49 between dominant and impulsive, raise some question as to the independence of such variables. While the theoretical importance of this is not pertinent here, it can be expected that the relationships that are present in the criteria scores will have an influence upon the correlations obtained between the graphic-motor movements and the temperament scores.

Table 3 contains the intercorrelation matrix of the movement variables. Of the 52 graphic-motor variables, ten are time scores and ten are pressure scores, one for each task. The question that now arises is: to what extent do the separate measures relate to each other, or does the task itself determine the manner of performance? In other words, are similar movements related?

1. *Time Variables:* Variables 2, 7, 12, 17, 22, 27, 32, 37, 42, and 48 are the time measures. It should be pointed out that each score was independently derived and the only obvious feature each one had in common was that each was a measure of the time it took to complete a particular task. Actually the time scores are measures of spontaneous or individual tempo, as no fixed time was set for any of the tasks nor were the subjects aware that their time was being recorded. As is indicated by Table 3, there is a large measure of agreement among the various time measures. All of the correlations are significant at the .05 level and most of them at the .01 level of significance. This suggests that there is a spontaneous tempo that comes out in all the tasks. The subjects' tempo was so consistent that, no matter what the task required, they performed it according to their preferred rate of speed. Thus, in the graphic movement tasks, individual tempo predominates and not the particular task. As might be expected, there is greater consistency in similar tasks than in dissimilar ones. Thus, the highest correlations were among the vertical line, horizontal line, and sagittal line. The lowest correlations were in the staircase task (Variables 22 and 27) where two distinct movements were required in doing the task. Apparently the shifting of directions (from

horizontal to vertical and vice versa) reduced the generality of an individual's tempo. By converting the intercorrelations of the time scores to z scores, it is possible to determine which time variable has the highest correlation with all the other time variables. Variable 7 was found to have an average z score of .77 with all the other variables, or an r of .65; this was the highest correlation between one variable and all the other time scores.

2. *Pressure Variables.* Variables 1, 6, 11, 16, 21, 26, 31, 36, 41, and 47 are pressure measures. As is indicated, there is a high degree of generality among the pressure variables, much greater than among the time scores. All the variables are related above the .01 level of significance. The range of correlation coefficients varies from .43 to .83. Thus, it seems that it makes little difference whether the subject makes a line up or down, horizontally or vertically—the pressure he applies is rather constant. Also, whether the task requires straight lines or interrupted movements, such as in the staircase tasks, does not matter very much. We continue to find that the correlations between similar tasks are the highest. For example, the correlation between ascending staircase pressure and descending staircase pressure was the highest value. In order to obtain one variable having the highest over-all correlation with all the other pressure variables, the same procedure was followed as in the time scores. Variable 6 was found to have an average z of .71 and an r of .61. Thus, it seems that, although the correlations among the pressure variables are not as low as among the time scores, the time variables have some higher correlations than the pressure variables. This would suggest that there is greater variability among the time scores and perhaps some of the

time scores might best be considered as somewhat distinct from other time scores.

3. *Length Variables.* The measurements of the three straight lines—the vertical (Variable 3), horizontal (Variable 8), and sagittal (Variable 13)—quite obviously provide length scores which may be related. In addition, the size of the circles (Variable 18) is a measure of length of diameter which may bear a relationship to the length of straight lines. In the staircase tasks, there are four length measurements, the vertical (Variables 24 and 29) and horizontal (Variables 25 and 30) lengths in both the ascending and descending staircase. It should be pointed out that the subjects were requested to reproduce the given length of the model staircase. Thus, just as in the straight lines and in the circle, there was an initial stimulus figure, but, unlike those two, the subjects were not required to trace over the stimulus figure. The ladders (Variables 45 and 51) also provide an additional source of length measurement. In every ladder task, the subject is required to reproduce the lines the same size as the stimulus lines. Thus, the average length of the lines provides a measure of the accuracy of reproduction of the length of the stimulus lines.

It is apparent from Table 3 that the over-all intercorrelations are not as high as were present among the time and pressure scores. Seven of the 45 intercorrelations were not significant and ten of them were significant between the .05 and the .01 level of significance. In other words, 28 of the 45 intercorrelations were significant at the .01 level and above. It is evident that this is well above chance. As might be expected, the highest correlations were among like variables. Variables 24, 25, 29, and 30—the staircase lengths—had the highest intercorrela-

TABLE 3—Continued

	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
-05	05	07	45	-03	14	06	08	48	-03	-05	07	13	54	04	11	12	03	-05	47	06	09	04	-00	18	1
16	05	04	-01	69	-02	06	04	02	43	15	-27	-20	-17	37	-09	-14	-30	16	-10	62	21	22	-25	01	2
35	38	34	11	18	28	-07	-14	09	25	35	-04	-05	16	33	22	20	25	-02	-10	22	06	12	34	-15	3
-24	-30	-24	08	-15	-18	04	07	-04	-18	-16	01	08	-06	-18	-04	01	-10	01	04	-22	-11	-19	-24	33	4
22	14	15	20	14	19	05	11	18	12	21	-16	-17	04	16	23	24	07	-08	01	12	23	16	12	17	5
04	12	10	46	01	22	-01	08	53	02	08	09	08	69	12	21	25	14	-11	57	08	17	15	10	12	6
22	03	01	02	71	-02	-12	-07	00	47	18	-32	-27	-20	41	-09	-16	-25	14	-12	62	16	16	-26	-04	7
15	24	23	35	18	25	-14	01	35	23	28	-07	-11	21	25	18	14	33	00	15	17	03	06	30	-25	8
11	01	05	08	-04	05	10	16	-02	00	09	-21	-18	-10	-04	05	10	05	01	-09	03	-00	09	-10	14	9
-10	05	-08	-10	-15	10	-01	-13	-08	-17	-17	-08	-05	10	04	32	34	-05	-05	13	-22	-05	-07	24	09	10
-18	-05	-01	43	-05	20	-03	05	48	-07	-04	13	22	53	02	18	18	19	-10	50	-05	02	-01	03	07	11
18	03	-03	05	64	-05	03	07	03	42	13	-30	-25	-17	36	-07	-13	-25	22	-09	59	12	13	-32	00	12
15	25	19	18	25	27	-15	-08	20	14	30	03	-02	12	23	30	20	32	-10	01	20	12	14	28	-24	13
22	14	14	11	06	15	09	-07	08	18	20	-11	10	15	10	11	21	-03	-08	04	12	30	34	21	21	14
08	-09	-06	04	15	11	13	19	-00	22	18	-13	-03	04	16	03	10	-07	30	10	19	16	25	-06	26	15
09	16	18	47	-06	18	14	-03	57	-05	06	-00	02	54	03	21	22	03	-10	55	05	24	24	07	27	16
22	06	08	-01	64	02	03	09	03	42	17	-29	-22	-17	35	-02	-09	-21	06	-16	57	24	26	-19	04	17
34	21	26	06	12	12	-08	-02	08	20	17	06	07	17	21	17	11	36	02	03	25	17	23	28	-03	18
08	16	24	02	-06	14	00	21	09	-08	-04	08	09	01	00	23	30	-14	02	01	-10	02	06	06	-02	19
-10	-20	-12	06	01	-03	00	13	03	-05	-03	-00	-05	02	07	10	13	01	-03	13	03	09	-02	-12	16	20
17	40	34	52	10	10	06	00	54	-01	-02	-08	-06	45	-02	-05	00	-15	08	48	07	04	-02	-05	27	21
38	18	20	03	44	25	-11	10	03	49	34	-22	-16	-12	41	17	17	10	04	-13	35	18	15	14	-14	22
77	55	60	06	05	23	01	06	09	16	26	-12	-10	11	29	24	21	16	-14	01	24	38	35	29	-04	23
53	78	72	14	04	23	-06	11	26	14	19	-12	-11	19	25	22	16	08	-14	14	12	23	20	22	-01	24
60	72	73	04	07	26	05	11	12	18	20	-11	-04	08	25	15	14	11	-15	07	18	26	26	25	-03	25
07	31	28	58	07	10	11	06	49	-03	-02	-10	-13	44	05	09	12	-19	05	43	00	02	-04	-08	23	26
47	28	31	03	41	18	-22	-04	02	40	32	-16	-09	-12	29	-03	05	07	-04	-21	37	17	14	18	-12	27
	60	67	14	20	19	03	03	09	25	26	-18	-11	07	29	03	01	13	-07	-04	38	39	35	25	05	28
		26	-11	28	-12	-08	33	17	21	-10	-08	28	21	12	-11	08	-12	28	18	27	19	26	08	29	
		27	-11	26	-05	-01	38	15	20	01	-04	22	21	11	06	14	-24	20	16	26	20	22	06	30	
			-27	30	07	23	76	16	20	-16	-19	64	16	24	23	16	-10	48	10	20	12	10	-02	31	
				38	-17	07	20	74	45	-30	-21	-01	66	17	12	-02	05	02	66	25	28	00	-05	32	
				-13	-04	30	40	63	-12	-07	23	50	49	41	25	28	17	14	22	28	33	-16	33		
					37	-03	-16	-14	-14	-02	-00	-24	-15	-10	-25	-23	02	00	03	04	-19	20	34		
						14	01	-11	-15	-06	08	-05	-03	07	-13	31	07	04	-10	-08	-29	06	35		
							06	11	02	-08	60	15	17	17	20	-19	53	12	21	15	11	-03	36		
								62	-29	-15	-01	78	21	15	13	-05	-06	64	33	35	18	-14	37		
									-19	-16	15	49	35	26	23	-21	-02	33	36	39	25	-14	38		
											11	-21	-01	03	16	-01	09	-22	-07	-04	-14	16	39		
											15	-12	09	-16	18	08	13	-01	15	19	-05	25	40		
														11	33	40	23	-12	68	01	20	18	17	02	41
															-54	46	18	-13	-04	62	34	32	18	-17	42
																	24	-25	26	02	28	30	-25	-15	43
																	15	-19	29	-03	-20	27	19	-03	44
																		-23	02	04	08	12	53	-28	45
																			-03	05	-30	-25	-40	24	46
																				-08	12	12	05	07	47
																					61	58	01	03	48
																						82	22	06	49
																							26	09	50
																								-33	51
																									52

Variable Number and Name

- | | |
|--|--|
| 27. Time-Descending Staircase | 40. Deviation-Downgoing Zigzags |
| 28. Number of Stairs-Descending Staircase | 41. Pressure-Upgoing Ladders |
| 29. Vertical Length-Descending Staircase | 42. Time-Upgoing Ladders |
| 30. Horizontal Length-Descending Staircase | 43. Number of Lines-Upgoing Ladders |
| 31. Pressure-Upgoing Zigzags | 44. Distance between Lines-Upgoing Ladders |
| 32. Time-Upgoing Zigzags | 45. Length-Upgoing Ladders |
| 33. Number of Angles-Upgoing Zigzags | 46. Deviation-Upgoing Ladders |
| 34. Distance Apart-Upgoing Zigzags | 47. Pressure-Downgoing Ladders |
| 35. Deviation-Upgoing Zigzags | 48. Time-Downgoing Ladders |
| 36. Pressure-Downgoing Zigzags | 49. Number of Lines-Downgoing Ladders |
| 37. Time-Downgoing Zigzags | 50. Distance between Lines-Downgoing Ladders |
| 38. Number of Angles-Downgoing Zigzags | 51. Length-Downgoing Ladders |
| 39. Distance Apart-Downgoing Zigzags | 52. Deviation-Downgoing Ladders |

tions. Variables 45, the length of the ladder going up, had the lowest correlation with all the other variables. It had only a slight positive relationship with the staircase lengths. It was found that variable 29 had the highest over-all correlation with all the other length measures; this correlation was .47.

4. *Intercorrelations among other Like Variables.* Most of the other variables fell into one of the four following categories: upward-downward, rightward-leftward, number of stairs, angles, lines, and deviation scores. Variables 4, 10, 14, and 20 are the four upward-downward variables. The only significant correlation was between Variables 4 and 20. This correlation was significant beyond the .01 level of significance. The other correlations are all slightly positive and the correlation between Variables 10 and 20 approach the .05 level of significance. It is not readily evident why upward shifting in the vertical line should be very significantly correlated with upward shifting in the circle. The circle was in a sagittal position and it might be expected that upward shifting in the sagittal line (Variable 14) would be more related to upward shifting in the circle. It may be that the rather low reliabilities of the upward-downward shiftings make these variables of questionable usefulness.

The intercorrelations of the four rightward-leftward variables (Variables 5, 9, 15, and 19) provide three significant relationships. The intercorrelations of the three line variables are all significant beyond the .01 level. It seems that rightward-leftward shiftings are not affected by the placement of the line—whether the line is parallel, perpendicular, or in a horizontal position, our subject shifted approximately to the same extent in all three spheres. It does appear, however, that the right-left shiftings in the circle

bear little relationship to the line shiftings. The correlation between the circle right-left shiftings is not so much affected by the position of the line as upward-downward shifting. This is reasonable, as in changing the position of the line, we are mainly rotating the vertical or the up-and-down dimension.

There are 15 correlations among the number of stairs (Variables 23 and 28), angles (Variables 33 and 38), and lines (Variables 43 and 49). Of these, seven are significant beyond the .01 level, five between the .05 and .01, and three are insignificant, although in the positive direction. It does not seem to make too much difference on what plane the movements are made or whether the direction is up or down; our subjects perform about the same. Again, it is found that there is a higher correlation among similar tasks than the not so similar ones. Thus, the highest correlations were between the ascending and descending staircase and between the zigzags going up and going down.

Another comparison was among the variables found in the zigzags (Variables 35 and 40) and ladders (Variables 46 and 52). Actually they are rightward-leftward shiftings, but because the task is much different from the lines and circle, it was felt that they were in no way comparable. The results indicate that there is not much consistency in this type of performance. Out of the six correlations, two are significant at the .05 level and one at the .01 level. It seems that the direction of the task, whether going up or down, is an important consideration. The ascending zigzags and the egocifugal ladder had the highest correlation (.31). There was a slightly negative relationship between the deviation scores on the up-going and down-going zigzags. There was a significant positive relationship be-

tween the egocifugal ladder, the down-going zigzag, and the egocipetal ladder. It, therefore, appears that, while direction may be important, it does not seem to make too much difference in the ladder task. In this area the results are inconclusive, as only three of the six correlations were significant.

Hypothesis 3: Certain types of graphic-motor movements go together and orderly relationships exist among the various movements.

The significant intercorrelations among the graphic-motor variables are presented in Table B³. Due to the excessive length of this table, it is not being presented in this section. This table not only lists the intercorrelations of the significant variables, but also provides the interpretation of the relationship. Correlations with like measurements are not included (i.e., pressure variables correlated with other pressure variables). A quick overview of the findings readily reveals the great consistency that is present. Almost all the pressure variables are significantly related to length measurements, the heavier the pressure the longer the length.

Time scores seem to break down into two categories: the lines and circle in one, and the more complex tasks—the staircase, zigzags, and ladders—in the other. Time scores in the first category are consistently related to short length, less distance apart in the zigzags, fewer angles, stairs, and number of lines in the ladders. Time scores in the second category are associated with long length, rightward movement, fewer angles, stairs,

and number of lines and greater distance between the lines in the ladders.

Length scores also seem to fall into two categories. In the first are all the length measures except for the ladders; in the second are the two length measures in the ladder tasks. In the first category we find downward movement, fast time, fewer stairs, angles, and number of lines, heavier pressure, longer length, and leftward deviation in the parallels and zigzags. In the second category are found slow time scores, fewer angles, few lines, less stairs, and leftward shifting in the ladders.

Upward shifting in the vertical line is consistently related to short movements and to greater number of stairs; in the horizontal line, upward shifting is related to short length, slow speed, and rightward movement. For the sagittal line, upward movement is associated with heavier pressure; rightward movement with fewer stairs, fewer lines, and greater distance between lines. In the circle, upward shifting is related to short length.

Rightward shifting in the vertical line is related to upward movement, fast time, fewer stairs, and fewer lines in the ladders. For the horizontal line, rightward shiftings are significantly related to other rightward movements. Rightward movements in the sagittal line are related to other rightward movements, both line and ladders, upward in the sagittal line, and fast time. In the circle, rightward movement is related to fast time, long lines in the staircase, and fewer number of lines in the ladders. Rightward deviations in the zigzags and ladders are significantly related to heavy pressure, slow time, and short length.

Fewer stairs in the staircase tasks are consistently related to long length, fast time, fewer angles, fewer lines, greater distance between the lines, upward move-

³ A 16-page table listing the significant intercorrelations of the graphic movement variables and temperament factors has been deposited with the American Documentation Institute. Order Document No. 5755, remitting \$2.50 for 35-mm. microfilm or \$1.75 for 6 by 8 in. photocopies.

TABLE 4
NUMBER OF SIGNIFICANT CORRELATIONS BETWEEN THE GRAPHIC-MOTOR VARIABLES
(.22 and above) (+ positive correlations, - negative correlations)

Variables	Pressure	Length	Ladder Length	Time in Lines & Circle	Time in Complex Tasks	Upward-Downward
No. of Variables in each group	10	8	2	4	6	4
Pressure	10					
Length	8	22+	1+	1-	1+	1+
Ladder Length	2	1+	12+		21+	10-
Time in Lines & Circle	4	1-		6-		1+, 1-
Time in Complex Tasks	6	1+	6-		24+	2-
Upward-Downward	4	1+		24+		2-
Rightward-Leftward	4		1+, 1-	2-	2-	
Deviation	2	2+			5+	2+
Number of Stairs	2	4+	3-	3+, 1-		1+
Number of Angles	2		5-	2+	9+	2+, 2-
Number of Lines	2	4+	2+		10+	
Distance between Lines	2	4+	4+		5+	2+
Distance Apart	2	6+	3+	2+	5+	2+
		2+	1+	2+	5+	
		1+	1-	4-	6-	

ment in the sagittal line, and downward and rightward movement in the vertical lines.

Fewer angles in the zigzags are associated with long length, heavier pressure, fast time, fewer stairs, fewer lines, and greater distance between the lines.

Fewer lines in the ladders are significantly related to long length, upward and rightward movements in the lines, fewer stairs, heavier pressure, leftward movement in the ladders, and greater distance between the lines.

Large distances between the lines on the ladders are associated with rightward and upward movements, heavier pressure, fast time, fewer angles, fewer stairs, and fewer lines.

Larger distances apart in the zigzags are significantly related to slow time, rightward deviations in the zigzags, and short length.

Table 4 summarizes the above findings. Under each category is listed the number of variables that fall within that category. Thus, for example, under length there are eight variables and under pressure, ten. Eighty significant

intercorrelations are possible; of these, 22 actually were significant. The sign behind the number of correlations identifies whether the correlations are positive or negative. In some cases there are significant relationships in both directions. The 13 categories were established on the basis of common relationships. For example, it was found that length in the ladders had significant relationships with variables which the other length scores did not have and vice versa. It should again be pointed out that chance would allow only one out of 20. It seems evident that something other than chance is operating in many of these relationships.

Hypothesis 4: Specific relationships between graphic-motor movements and temperament factors are postulated.

Table 5 presents the correlations between the graphic-motor variables and the seven temperament factors. As is evident, the correlations are quite low, with the highest in the thirties. For this table, a one-tailed test was used and an r of .18 is significant at the .05 level and

TABLE 4—*Continued*

Variables		Right- ward- Left- ward	Devia- tion in Complex Tasks	No. of Stairs	No. of Angles	No. of Lines	Dis- tance between Lines	Dis- tance Apart
No. of Variables in each group		4	4	2	2	2	2	2
Pressure	10		4+		4+	4+	6+	1+
Length	8	2+	3-	14+	10+	7+	2+	
Ladder Length	2		5-	2+	4+	3+	1+	1-
Time in Lines & Circle	4		3+, 1-	2+		2+	2+	4-
Time in Complex Tasks	6	5+		9+	10+	5+	5+	6-
Upward-Downward	4	2+	1+	2+, 2-		2+	2+	
Rightward-Leftward	4		2+	2+		3+	3+	
Deviation	4	2+			1-	2-	1-	2+
No. of Stairs	2	2+			3+	3+	3+	
No. of Angles	2		1-	3+		4+	4+	
No. of Lines	2	3+	2-	3+	4+		3+	
Distance between lines	2	3+	1-	3+	4+	3+		
Distance Apart	2		2+					

an r of .26 at the .01 level. Due to the extremely low correlations obtained, the findings will be examined to determine if any trend exists, with the understanding that any small but significant correlations must be accepted very tentatively and that further research on a larger sample is needed before any firm conclusions are possible.

(a) *High scorers on the Active Factor will work faster and exert heavier pressure.* The correlations between the time scores and the Active Factor reveal four significant relationships. At the .01 level, time in the horizontal (Variable 7) and in the sagittal (Variable 12) lines is significantly related, and at the .05 level, time in the circle (Variable 17) and in the upgoing zigzag (Variable 32) has significant associations. Of the other six, four are in the predicted direction and two are negative. The correlations of the pressure variables and the Active Factor reveal no significant relationships; however, nine out of the ten coefficients fall in the predicted direction.

(b) *High scorers on the Vigorous Fac-*

tor will be found to have heavier pressure, fewer parallel lines, and faster time scores. There are three significant correlations (Variables 1, 11, and 16) at the .05 level between the pressure variables and the Vigorous Factor. Five of the other correlations are in the predicted direction and two—the ascending staircase and the upgoing ladders—were slightly negative. It may be that the vigorous person does not exert the same force when he has to draw away from himself.

Fewness of the lines in the parallel or ladder task was measured by the number of lines it took to get to a certain distance and the average distance between the lines. As is indicated in Table 5, the correlations were insignificant in the upgoing ladder (Variables 43 and 44) but significant in the downgoing ladder (Variables 49 and 50) at the .05 level. Thus, the trend seems to be that direction is important and, just as the vigorous person exerts more pressure when the task requires downward movements, so, too, there are fewer lines in covering a

TABLE 5
CROSS-CORRELATIONS OF MOVEMENT VARIABLES WITH TEMPERAMENT FACTORS

Variable Nos.	Active 53	Vigorous 54	Impulsive 55	Dominant 56	Stable 57	Sociable 58	Reflective 59
1	10	18	06	12	-07	10	-03
2	16	17	16	-08	22	10	-10
3	-01	-08	-11	04	-02	-12	34
4	-20	-12	04	-03	00	09	01
5	05	-35	-12	11	-07	04	06
6	08	11	03	12	-21	10	03
7	22	16	12	-03	26	05	-07
8	02	09	07	00	-06	-07	10
9	-14	-29	-02	13	-02	18	02
10	07	-10	-08	17	01	21	16
11	00	21	16	14	-20	05	10
12	20	08	11	-08	25	04	-09
13	14	03	05	-03	-01	-14	13
14	07	-13	-11	-12	02	-13	04
15	-05	-21	00	02	14	12	00
16	08	18	01	02	-06	-02	11
17	18	09	09	-05	29	03	-23
18	-10	-03	-19	-07	-06	-12	27
19	06	09	-04	-11	-09	-07	02
20	-06	-06	-22	-19	06	-03	-19
21	10	06	01	03	-01	-05	-03
22	-01	06	02	18	-05	-13	-03
23	02	01	-21	-10	-04	-15	03
24	-08	-02	-24	-05	-08	-19	09
25	-07	05	-20	-14	-07	-18	10
26	15	06	-05	08	02	-02	00
27	-02	08	05	-17	04	-08	00
28	03	-06	-18	-17	-07	-24	08
29	-12	-04	-27	-08	-02	-15	21
30	-01	-01	-20	-12	-03	-22	22
31	10	06	09	15	13	-03	09
32	19	13	17	12	21	11	-03
33	04	10	13	24	10	11	-06
34	-03	01	01	-03	24	-04	01
35	-03	05	02	-03	00	-11	-12
36	02	08	08	08	00	-09	06
37	01	21	14	04	19	02	-01
38	08	02	10	11	27	11	-12
39	-12	-10	-09	-17	-15	-16	15
40	-19	-03	-07	-14	-01	-13	13
41	00	-05	-04	10	-02	01	21
42	14	18	03	04	12	-01	04
43	17	11	06	16	04	10	06
44	09	-06	-06	10	-01	06	14
45	-01	-04	01	09	-13	-06	05
46	-06	-09	-07	-06	-02	-07	00
47	-07	05	-03	12	-01	09	12
48	07	23	-01	-06	21	00	-06
49	02	24	-02	-02	27	11	-15
50	02	25	-03	-01	22	06	-11
51	03	-01	01	13	02	-02	04
52	-06	-20	-18	05	04	-01	06

$r = .18$ for the .95 level and .26 for the .01 level.

distance when he goes downward or toward himself.

All the correlations between the Vigorous Factor and the time scores are in the predicted direction, with Variables 37,

42, and 48 significant at the .05 level. The trend appears to be that there is some relationship between speed and the Vigorous Factor.

(c) High scorers on the Impulsive Fac-

tor will shift sooner, work faster, and have greater pressure. Utilizing the staircase task as an indication of ability to shift quickly and the length measurement as a measure of this, and based upon the prediction that those who shift quickly have the shortest average length on both the vertical and horizontal length, it was found that there is a significant negative relationship between the Impulsive Factor and length on these tasks. Variables 24, 25, and 30 are significantly related to the Impulsive Factor at the .05 level and Variable 29 at the .01 level. It appears that the subjects who score high on impulsiveness have the shortest length on these tasks and, therefore, shift the quickest.

The correlation of the time variables with the Impulsive Factor reveals that, in nine of the ten cases, the relationship is in the predicted direction. No significant correlations were obtained. Similarly, the correlations of the pressure scores with the Impulsive Factor provided no significant relationships with seven of the ten correlations in the predicted direction. Thus, for both time and pressure scores, the trend seems to agree with the hypothesis, but the correlations are so low that no definitive conclusions can be made.

(d) *High scorers on the Dominant Factor will shift upward, have larger movements in reaching a certain distance, and will exert more pressure.* Of the four measures of upward shifting—the vertical line, horizontal line, sagittal line, and circle—only the upward shifting in the horizontal line is related to the Dominant Factor in the predicted direction. This would seem to indicate that upward shifting as such is not related to dominance. The horizontal shifting is different from the other three tasks in that the subject is required to reproduce lines from right

to left and vice versa. Thus, there is less likelihood of up and down shifting. It may be that, when the task requires upward or downward movement, not enough individual expression is brought forth, and when left or right movement is called for, upward or downward shifting would be more revealing of temperamental differences.

In order to determine whether dominant people have larger movements in reaching a certain distance, the number of angles and lines it took to go from one part of the page to the other side were correlated with the Dominant Factor. In the upgoing direction, there was a significant relationship (at the .05 level) between fewness of angles and dominance; the relationship between fewness of lines was not quite significant but approaches significance at the .05 level. For the downgoing angles, there was a slight positive correlation and for the downgoing lines, there was a slight negative correlation. It therefore seems that the trend suggests some relationship between large movements and the Dominant Factor in the upgoing direction, but questionable relationship in the downgoing direction.

All the correlations between the pressure variables and the Dominant Factor were found to be positive, although none of them is significantly related to this factor. Thus, there is some suggestion of a relationship between dominance and pressure.

(e) *High scorers on the Stable Factor will shift in the upward direction, work faster, and will make more outgoing movements.* As is indicated in Table 5, there were no significant correlations between upward movement (Variables 4, 10, 14, and 20) and the Stable Factor. All the correlations are in the positive directions, but very low. It therefore appears

that, while there may be some relationship between upward shifting and this factor, the correlations are too small to draw any conclusions about such relationships.

On the other hand, the correlations between the time variables and this factor provide many significant associations. Except for Variable 22, all the correlations are positive, with two significant variables (7 and 17) at the .01 level and five at the .05 level (Variables 2, 12, 32, 37, and 48). The trend is consistent with the prediction that a relationship exists between fast tempo and the Stable Factor.

Utilizing the number of lines and angles as a measure of outgoingness, it was found that all four correlations are in the predicted direction. The two downgoing variables are significant beyond the .01 level, while the two upgoing variables are low positive correlations. These findings suggest a relationship between high Stable scores and large movements in the downward direction. This is the direct opposite of the finding in relation to the Dominant Factor. This raises the question why do "Stable" people utilize fewer movements in the downward direction while "Dominant" people utilize fewer movements in the upward direction?

(f) *High scorers on the Sociable Factor will shift in the rightward and upward direction.* The correlations between the rightward-leftward variable and the Sociable Factor reveal that three of the four coefficients are in the predicted direction and one slightly negative. The only significant correlation was in the horizontal line and this was significant at the .05 level. For the upward-downward shiftings there were two correlations in the predicted direction and two were negative. Again the only significant corre-

lation was the shifting in the horizontal line. Thus, it seems that the results do not confirm the hypothesis concerning the relationship between rightward and upward movement and the Sociable Factor. There is some indication that, in the horizontal line, subjects who score high on the Sociable Factor shift significantly more rightward and upward.

(g) *High scorers on the Reflective Factor will work more slowly, shift more slowly, and will have leftward or inward movements.* Nine of the ten time variables are negatively related to the Reflective Factor, suggesting that slow time is associated with this factor. The only significant correlation was Variable 17 (circle) at the .05 level. Thus, there is some confirmation of the hypothesis that high scorers on the Reflective Factor have slow natural tempo.

Although only the length variables in the staircase task were used to measure ability to shift quickly in relation to the Impulsive Factor, the Reflective Factor has been compared with all the length variables. While the tendency to persevere may best be revealed in a task in which quick shifting is necessary, it is felt that the tendency to continue a task may also be indicated in any short task where change of direction is necessary. The correlations between the Reflective Factor and the length variables reveal that nine of the ten correlations are positively correlated, suggesting that those who score high on this factor draw longer lines. Two of the correlations are significant beyond the .01 level (Variables 3 and 18), and two at the .05 level (Variables 29 and 30). Thus, there is some confirmation that reflective people cannot "let go" and have a tendency to continue a graphic movement once they have started.

It was hypothesized that leftward

movement would be correlated with the Reflective Factor. This would mean negative correlations with rightward-leftward shifting variables. The four correlations between these variables and the Reflective Factor reveal that only one of the four was in the predicted direction and, in all cases, the correlation was very slight. It therefore seems that this hypothesis was not upheld and that the data do not indicate any relationship between leftward movement and the Reflective Factor.

DISCUSSION

The present investigation had the purpose of studying expressive graphic-motor movements in order to discover if they represent an orderly, consistent, and meaningful domain of personality. To determine this, four major hypotheses were formulated. These were: graphic-motor movements are stable measures and therefore reliable; individuals are consistent in their performance on a variety of graphic-motor tasks measuring similar kinds of movement; an orderly relationship exists among the movements; and graphic-motor movements can be revealing of temperamental differences. Thus, reliability was studied; then intraconsistency, generality, and, finally, validity.

In studying reliability, various problems arise. The question of a learning effect is always to be considered in test-retest reliability. Therefore, considerable time is required between original testing and retesting in order to insure that the performing of the original test is not very well remembered. Whether learning takes place on the Mira technique appears questionable. It does not seem that much learning can take place on a technique where no verbal responses are given and the subject is unable to see what he does. Thus, although a two-week period may seem of small duration for test-retest reliability, it is felt that this is sufficient time for a task of this nature.

It is well known that self-correlations are affected by the range of test variables and the

population sampled. There is little question that the present sample is a very select group in that only college students were used. There is no way of knowing what the effect has been on the reliabilities because of this select sample. Much higher reliabilities may have been obtained if a wider range of normal and perhaps abnormal subjects were utilized. While it has never been claimed that college students all have the same temperament, it still appears that most students are more similar to each other than to other groups. It may, therefore, appear that if a more diffuse group were used, the reliability would have been even greater.

An average reliability coefficient of .617 was obtained. Is this sufficiently high so that it can be said that these subjects are stable in their graphic-motor movements? Allport & Vernon (4) obtained an average coefficient of .64 in different sessions; Eisenberg (18) reported a coefficient of .76. Compared with other findings, the present results are in substantial agreement. Based upon the previous work done on expressive movements, it seems that we are on solid ground when we state that, on the average, expressive graphic-motor movements are fairly stable characteristics in our subjects.

Next comes the problem of intraconsistency. We have attempted to discover if similar variables are related in graphic-motor performance. Allport & Vernon (4) found three speed "factors," verbal, drawing, and rhythmic; and they state, "While there is no conclusive evidence of a single, 'pervasive, 'personal tempo' it is false to conclude that there is a specific speed for each activity and for each occasion." It is apparent that the domain covered in the present study is much more limited than Allport & Vernon's. We have only measures of graphic-motor time, how long it took a subject to complete a task when no mention was made of time. The results indicate that there is a large degree of consistency in the time scores. It was also found that other factors operate to reduce the correlation between time scores. One striking illustration was on the staircase task where directional changes were required. People who worked fast on this task usually had longer length, as shifting made for longer time. In other tasks, fast time usually meant short length, as might be expected. Therefore, it is concluded that there is a consistency in time scores on the graphic-motor tasks; various factors may interfere with this consistency but basically it still exists.

In addition to the high degree of intraconsistency found among the time variables, high intercorrelations were obtained among pressure and length variables. There was lesser agreement among the rightward-leftward variable and the number of angles, lines, and stairs; yet, the number of significant relationships was well

above chance and indicate that among like tasks the subjects are fairly consistent. It was only in the upward-downward variables and in the deviations measures of the zigzags and ladders that the results are not so clear-cut; however, even here the findings are above what would be expected by chance alone.

Earlier it was stated that one of the purposes was that of contributing to understanding of the unity of personality through graphic-motor movements. Actually, each of the four hypotheses represents an attempt to do just that. If people are consistent over a period of time in their movements, there is unity; if subjects perform the same type of movement similarly, there is unity; if there is a meaningful relationship between various types of movements, there is unity; and if one domain that is so vastly different in kind of material elicited is found to be related to another domain, i.e., graphic-motor movements and temperament scores on a questionnaire, there is also unity.

The first two hypotheses have already been discussed; the third hypothesis, that of generality, is fairly well upheld. That is, the relationships between different kinds of variables are generally consistent. For example, out of the 80 correlations between pressure and length, 22 are significant; out of the 24 correlations between time in lines and circles and length, 21 are significant. Of the 16 correlations between length and number of angles, 10 are significant. Of course, some of the most outstanding examples have been selected and it may be claimed that some of the results would naturally be expected, such as those who complete a task rapidly would have short lines; if they made longer lines, they would take longer time. This may or may not be so. We did not know this before we started. Allport & Vernon (4) reported that pressure and distance between lines was very much related. This would not commonly be suspected and perhaps was only due to the special conditions of their experiment. However, it was also found that of the 12 correlations between pressure and distance between lines, six were significant in the predicted direction.

We now come to the final hypothesis, that of the relationship between movement variables and temperament. In the final analysis, if graphic-movement variables are to be utilized, it is first necessary to obtain some indication of validity. How well do they correlate with other measures? This was a relatively uncharted area in which few objective investigations have been conducted. Specific hypotheses were formulated based upon a few researches and many unsubstantiated claims. Thus, they were all quite tentative and exploratory.

In a study of this sort and in most studies where another instrument is used as a criterion,

various problems arise. Rosenzweig (44) suggests the crucial importance of using criteria that tap the same level of behavior as the instrument to be validated. According to Rosenzweig, there are three levels which are tapped by psychodiagnostic tests. The first is that of self-critical or censored responses from which one may predict towards other situations in which the subject is self-consciously critical. Level II is that of overt behavior; if the subject manifests in the test situation a certain type of overt behavior, one may predict that he will behave similarly in similar externally defined situations. Level III is that of implicit behavior, which requires interpretation from the manifest content of the response to underlying factors.

Expressive movements have often (6) been lumped together with projective tests as implicit behavior testing. The material itself does not indicate the manner of interpretation and the subject is totally unaware how his responses will be evaluated. If we are dealing with Level III behavior in graphic-motor tasks, can we fruitfully use as a criterion responses obtained by a temperament schedule? The Thurstone Temperament Schedule is somewhat different from most tests of this sort in that, first, the factors were obtained through a systematic factor analysis, and second, the original items were refactored and a second factor analysis was performed. As a result, the factors obtained are second-order factors. Baehr (8) has claimed that second-order factors represent a purification process and yield factors which represent the more basic or pervasive characteristics of the original behavior items included in the factorial study. In other words, she claims that by a second-order factor analysis the more basic aspects of the individual's temperament are obtained. If this is so, then there may be an equivalent Level III behavior resulting from the Thurstone Temperament Schedule.

As Cronbach (12) points out, the normal personality presents troublesome problems of measurement. While it is easy to identify abnormals since they tend to fall into distinct syndromes or patterns, intermediate positions, falling in the center of the distribution, tell the investigator little except that the subject is not a deviate. This suggests that individual differences are difficult to detect in the normal range, and there is always the question whether the scale is sufficiently refined to detect the differences that exist.

In dealing with a test in which the subject has to give answers about his own personality, there is always the question of truthfulness, and secondly, of knowledge. We never know how honest the subject is in reporting about himself. No matter what explanations are given and how often he is told that the results will have no

influence on any of his activities, one is never sure that the subject is able to respond in an honest fashion. Allport (3) states, "The normal subjects tell you by the direct method precisely what they tell you by the projective method. They are all of a piece. You may therefore take their motivational statements at their face value, for even if you probe, you will not find anything substantially different" (p. 110). Davids (13) compared normals on direct and indirect tests and found that normals have a significant degree of agreement among rank orders when guaranteed anonymity, but when they were told that the results would be used for job assessment, there was a much lower agreement.

On the basis of the above experiment, it appears that if it is certain that the subjects believe that the results will not be used in any way against them, the results may validly be compared with projective tests. In the present study we cannot be sure that such was the case. We can only state that the conditions were so set up that constant assurance was given to the subjects that the results would in no way affect them.

As for the second question, there is really no way of telling how much the subject knows about himself. While he may have a certain picture of himself, others may have quite a different picture. There is little question but that the information the subject provides is either his open self-picture or his censored self-picture, censored to show him in the most favorable light. It would be necessary to have others either rate him or attempt to take the test the way they believe the subject would answer the questions. This we did not do and therefore have to depend upon the answers the subject gave of himself as being the best picture obtainable under the circumstances.

The question may be raised—why was not some other third level test given that gets at implicit behavior as a more adequate criterion measure? Considerable thought was given to the selection of a criterion test, and it was felt that if we utilized a projective test, the problem of classifying the responses would require many assumptions that might be quite arbitrary. In addition, it should be emphasized that we were dealing with a normal population and the projective tests are mainly oriented toward abnormality. It would, however, be fruitful to test out the certain postulated comparisons, such as rightward movement on the Mira technique would be related to a predominance of color over movement responses on the Rorschach. This is an area for another study and was not the concern of the present investigation.

The correlations between the graphic-motor variables and the temperament factors were all very small, only a few above .30. Thus, in gen-

eral, the findings are extremely limited and the positive results that were obtained must be considered very tentative; further research on a larger sample is needed before any firm conclusions can be made. It is revealing that some of the results seem promising and suggest the possibility that an objective test of temperament may be devised utilizing graphic-motor movements. Other measures may be more useful for this purpose. It is necessary to have a beginning point and, although the present findings are very limited and in many instances inconclusive, this study is an attempt to further increase investigation and interest in this relatively uncharted avenue.

In concluding this discussion, it should be pointed out that this investigation is only the first step in exploring the area of graphic-motor movements. The findings indicate that the area of graphic-motor movements is a profitable one for psychologists to explore. At the present time generalizations cannot be made, and specific studies will have to be performed in order to test out possible similarities between the present findings and claims made by drawing analysts and graphologists. We utilized a specific test and obtained graphic-motor movements elicited by this test; whether the same results will be found by other techniques will have to be determined by investigation. It does seem, on the basis of the present findings, that tests of graphic-motor functioning can be devised which may be utilized in a diagnostic manner so that objective tests of temperament may be realized. Additional studies must be performed utilizing groups of children, more heterogeneous normal subjects, and abnormal subjects in order to substantiate the present results and to gain a greater degree of understanding in this complex area of personality.

SUMMARY AND CONCLUSIONS

A number of studies in the area of expressive movements have indicated that this is a meaningful avenue of exploration and have yielded positive results, revealing the intraconsistency of human functioning, both in the movement aspect and other characteristics of human behavior. However, previous studies have not adequately handled the three major interrelated aspects of the problem of expressive movement—the reliability of expression, the extent of unity or generality in expressive movement, and the relationships of expressive movement

measures to other behavior characteristics.

The present investigation deals only with graphic-motor expression, attempting to test the three interrelated aspects by objective means. Four major hypotheses were formulated:

1. Graphic-motor movements are stable measures, and test-retest correlations will reveal a high degree of reliability.

2. Individuals are internally consistent in their graphic-motor movements. Similar movements will be highly correlated.

3. Certain types of graphic-motor movements go together and orderly relationship exists among the various movements.

4. Certain specific relationships between graphic-motor movements and temperament factors were hypothesized.

Eighty male college students were administered a modification of the Mira Myokinetic Technique in which graphic movements are reproduced while the subject's vision is screened. Ten separate tasks were given and many of them contained similar variables such as tempo, pressure, and length measurements. The Thurstone Temperament Schedule was also administered in order to obtain a measure of temperament differences.

The approach throughout was a correlational one and 1,891 correlations were obtained. The first hypothesis was partially confirmed by a test-retest reliability coefficient of .62. The second hypothesis was largely confirmed, although all the results were not equally significant. The findings indicate that individuals are consistent in a number of important variables such as personal tempo, amount of pressure applied, and length of graphic reproductions.

The third hypothesis was also partially confirmed and meaningful consistent significant correlations were obtained

among different types of movement variables. The nature of the relationships became evident only after the data were analyzed. For example, significant consistent relationships were obtained between pressure and length variables, time and length variables, and between length and downward shifting variables.

The fourth hypothesis was tested by examining hypothesized relationships between the temperament factors and the movement variables. The correlations obtained were extremely low with only a few as high as .30. There were a number of significant correlations but it was apparent that the findings were very tentative and further research on a larger sample is needed before any firm conclusions are possible in relating graphic movements with temperament. The trend of the findings revealed: the Active Factor had positive relationships with time and pressure scores; the Vigorous Factor had positive associations with pressure, fewness of lines and tempo; the Impulsive Factor is negatively related to the length scores where shifting was required and had positive relationships with tempo and pressure scores; the Dominant Factor related positively with the pressure variables, fewness of angles, and fewness of parallel lines; the Stable Factor had significant positive correlations with most of the time variables and the number of angles and lines it took to reach a certain distance also had significant relations with this factor; the Sociable Factor had positive correlations with upward and rightward shifting variables; the Reflective Factor had negative correlations with the time variables and positive correlations with the length variables when shifting was required.

It was concluded that the present re-

sults seem quite promising in revealing the basic unity of an individual's graphic-motor movements and indicate that this area is a fruitful one for further investi-

gation and may be utilized profitably in establishing objective measures of temperament.

REFERENCES

1. ALLPORT, G. W. An experimental study of the traits of personality. Unpublished doctoral dissertation, Harvard College Library, 1922.
2. ALLPORT, G. W. *Personality: A psychological interpretation*. New York: Holt, 1937.
3. ALLPORT, G. W. The trend in motivational theory. *Amer. J. Orthopsychiat.*, 1953, 23, 107-119.
4. ALLPORT, G. W., & VERNON, P. E. *Studies in expressive movements*. New York: Macmillan, 1933.
5. AMBREDANE, A., ANGELIN-SHUTZENBERGER, A. E., & FAVERGE, J. Sur la fidelite du test myocinetiques of Mira y Lopez. In F. Baumgarten (Ed.), *La psychotechnique dans le monde moderne*. Paris: Presses Univer. de France, 1952. pp. 180-181.
6. ANDERSON, H. H., & ANDERSON, G. L. *Introduction to projective techniques*. New York: Prentice-Hall, 1951.
7. ARNHEIM, R. Experimentall-psychologische untersuchungen zum ausdrucksproblem. *Psychol. Forsch.*, 1928, 11, 1-132.
8. BAEHR, M. E. A factorial study of temperament. *Psychometrika*, 1952, 17, 107-126.
9. BELL, J. E. *Projective techniques, a dynamic approach to the study of personality*. New York: Longmans, Green, 1948.
10. BLOOR, C. *Temperament*. London: Methuen, 1928.
11. BÜHLER, C. The social behavior of the child. In C. Murchison (Ed.), *Handbook of child psychology*. Worcester, Mass.: Clark Univer. Press, 1931. pp. 392-43.
12. CRONBACH, L. J. *Essentials of psychological testing*. New York: Harper, 1949.
13. DAVIDS, A. Comparison of three methods of personality assessment: Direct, indirect and projective. *J. Pers.*, 1955, 23, 423-440.
14. DEUTSCH, F. Analysis of postural behavior. *Psychoanal. Quart.*, 1947, 16, 195-213.
15. DOWNEY, J. E. *Graphology and the psychology of handwriting*. Baltimore: Warwick & York, 1919.
16. DOWNEY, J. E. *The will-temperament and its testing*. Yonkers: World Book, 1923.
17. EISENBERG, P. Expressive movements related to feelings of dominance. *Arch. Psychol.*, 1937, No. 211.
18. EISENBERG, P. A further study in expressive movement. *Charact. & Pers.*, 1937, 5, 296-301.
19. EISENBERG, P. Judging expressive movements: I. Judgments of sex and dominance feeling from handwriting samples of dominant and nondominant men and women. *J. appl. Psychol.*, 1938, 22, 480-486.
20. EISENBERG, P. The motivations of expressive movement. *J. gen. Psychol.*, 1940, 23, 89-101.
21. ENKE, W. Die psychomotorik der konstitutionstypen. *Z. sch. f. ang. Psychol.*, 1930, 36, 237-287.
22. GESELL, A. *An atlas of infant behavior: A systematic delineation of the forms and early growth of human behavior patterns*. New Haven, Conn.: Yale Univer. Press, 1934.
23. GUILFORD, J. P., & GUILFORD, R. B. Personality Factors D, R, T and A. *J. abnorm. soc. Psychol.*, 1939, 34, 21-36.
24. GUILFORD, J. P., & GUILFORD, R. B. Personality Factors N and GD. *J. abnorm. soc. Psychol.*, 1939, 34, 239-348.
25. HARMS, E. The psychology of formal creativeness. I. Six fundamental types of formal expressions. *J. genet. Psychol.*, 1946, 69, 97-120.
26. JACOBY, H. J. The handwriting of depressed children. *New Era*, 1944, 24, 4-6.
27. JOHNSON, B. Changes in muscular tension in coordinated hand movements. *J. exp. Psychol.*, 1928, 11, 329-341.
28. KEMPF, E. J. *The autonomic functions and the personality*. New York: Nervous & Mental Disease Publ., 1921.
29. LANDIS, C. The interpretation of facial expression in emotion. *J. gen. Psychol.*, 1929, 2, 59-72.
30. LEVINE, A. S. Preservation, rigidity and persistence. *Psychol. Rep.*, 1955, 1, 107-125.
31. LUNDHOLM, H. The affective tone of lines: Experimental researches. *Psychol. Rev.*, 1921, 28, 43-60.
32. LURIA, A. R. *The nature of human conflicts*. New York: Liveright, 1932.
33. MASLOW, A. H. *Motivation & personality*. New York: Harper, 1954.
34. MIRA, E. Myokinetic psychodiagnosis: A new technique for exploring the conative trends of personality. *Proc. roy. Soc. Med.*, 1940, 33, 173-194.
35. MIRA, E. *Psychiatry in war*. New York: Norton, 1943.

36. MIRA, E. Etude sur la validite du test psychodiagnostique myokinetique. *Annee psychol.*, 1951, 50, 575-584.
37. MIRA, E., MIRA, A., & ALIVEIRA, A. Aplicacao do psicodiagnostico miocinetico ao estudo da agressividade. *Arch. brasil Psicotecnica*, 1949, 1, 69-16.
38. MUNSTERBERG, H. Die psychophysische grundlage der gefuhle. *Proc. II. Int. Congr. of exp. Psychol.* London: William and Norgate, 1892, 132-135.
39. PASCAL, G. R. Handwriting pressure: Its measurements and significance. *Charact. & Pers.*, 1943, 11, 234-354.
40. PERL, W. R. On the psychodiagnostic value of handwriting analysis. *Amer. J. Psychiat.*, 1955, 111, 595-602.
41. POFFENBERGER, A. T., & BARROW, B. E. The feeling value of lines. *J. appl. Psychol.*, 1924, 8, 187-205.
42. RIBBLE, M. A. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944, 621-661.
43. ROMAN, K. G. *Handwriting*. New York: Pantheon, 1952.
44. ROSENZWEIG, S. Levels of behavior in psychodiagnosis with special reference to the picture frustration study. *Amer. J. Orthopsychiat.*, 1950, 20, 63-72.
45. SPITZ, R. The smiling response: A contribution to the ontogenesis of social relations. *Genet. Psychol. Monogr.*, 1946, 34, 57-125.
46. THURSTONE, L. L. *The factorial description of temperament*. Report No. 53, Psychometr. Lab., Chicago: Univ. Chicago Press, 1950.
47. THURSTONE, L. L. *Experimental tests of temperament*. Report No. 67, Psychometr. Lab., Chicago: Univer. Chicago Press, 1951.
48. THURSTONE, L. L. *The development of objective measures of temperament*. Psychometr. Lab., Chapel Hill: Univer. North Carolina, 1953.
49. WASHBURN, R. W. A study of smiling and laughing of infants in the first year of life. *Genet. Psychol. Monogr.*, 1929, 6, 397-535.
50. WANGER, M. A. An attempt to appraise individual differences in level of muscular tension. *J. exp. Psychol.*, 1943, 32, 213-225.
51. WHITEHORN, J. C. The material in the hands of the biochemist. *Amer. J. Psychiat.*, 1935, 92, 314-318.
52. WOLFF, W. Projective methods for personality analysis of expressive behavior in preschool children. *Charact. & Pers.*, 1942, 10, 309-330.
53. WOLFF, W. *The expression of personality*. New York: Harper, 1943.
54. WOLFF, W. *The personality of the pre-school child*. New York: Grune & Stratton, 1946.
55. WOLFF, W. *Diagrams of the unconscious*. New York: Grune & Stratton, 1948.
56. WOODWORTH, R. S., & SCHOLSBURG, H. *Experimental Psychology*. (2nd Ed.) New York: Holt, 1954.
57. WULFECK, W. H. Motor function in the mentally disordered: I. A comparative investigation of motor function in psychotics, psychoneurotics and normals. *Psychol. Rec.*, 1941, 4, 271-323.
58. WULFECK, W. H. Motor function in the mentally disordered: II. The relation of muscle tension to the performance of motor tasks. *Psychol. Rec.*, 1941, 4, 326-348.
59. ZOEPFFEL, H. Ein versuch Zur experimentellen feststellung der personlichkeit im sanglingsalten. *Z. f. Psychol.*, 1929, 111, 273-306.

(Accepted for publication June 24, 1958)

Psychological Monographs: General and Applied

Age-Sex Roles and Personality in Middle Age:
A Thematic Apperception Study

BERNICE L. NEUGARTEN AND DAVID L. GUTMANN

Committee on Human Development, University of Chicago

THE ORIGINAL purpose of this investigation was to explore the use of projective techniques in the study of social roles, and, more specifically, to use the Thematic Apperception technique in studying adult age-sex roles in the family.

There has been little empirical research on marital or parental roles in middle age, just as there has been relatively little research in the wider area of psychological and social changes that occur in men and women in the middle years of life. The investigators hoped to find fruitful ways of describing age-sex roles in the family and thereby to contribute to the social psychology of middle age.¹

¹ The data being reported here were gathered in connection with a larger study known as the Kansas City Study of Adult Life. The larger study, supported by a grant from the Carnegie Foundation, includes data on over 700 adults aged 40 to 70 residing in the metropolitan area of Kansas City. It has been carried out under the auspices of the Committee on Human Development of the University of Chicago, and under the direction of a research committee of which Robert J. Havighurst has been chairman.

The present study was supported in part, by funds provided by the Social Science Research Committee of the University of Chicago. The investigators are indebted to Robert Yufit for his assistance in pretesting the picture upon which this study is based; to Warren A. Peterson, who directed the field work involved in gathering these data; to Meyer Braiterman and Gunther Rice for checking the consistency of responses obtained from this picture with those obtained from standard TAT pictures; and to

The concept of social role involves not only the behavior exhibited by persons who occupy a common 'office' or status, but also the expectations held by other people about what that behavior should be. Expectations of role behavior are, in turn, both explicit and implicit. The investigator who is interested in what husbands expect of wives may choose to ask husbands explicitly by using interviews or questionnaires. He may, on the other hand, choose to deal with implicit expectations by exploring husbands' fantasies about wives.

At least two considerations prompted the choice of the Thematic Apperception technique. The first was that the respondent would be providing data about family roles without an explicit awareness that he was doing so. As a result, the responses should be relatively uncensored, more closely related to the respondent's personal values and experiences than those he might feel constrained to give in answer to more direct questions. Second, fantasy material, while presenting certain difficulties of analysis as compared with questionnaire data, should have a decided advantage for exploratory research. The richness and unstructured nature of projective data enable the investigator to follow an inductive process: he can follow up clues

Lisa Cohen for her assistance in establishing the reliability of our judgments.

as they appear in his data rather than check dimensions and hypotheses defined in advance.

The Thematic Apperception Test and adaptations thereof emerged from the field of clinical psychology and have been used mainly in the study of individual personality. They have also been used, however, to study groups of people, as, for instance, in studies of personality patterns common to persons of different ages or different cultures (1, 2).

While the study being reported here is of the latter type, it constitutes a departure in method insofar as attention was focussed, not upon the persons who supplied the data, but upon the persons about whom data were given. The primary concern was with the collective role-images of husbands, wives, sons, and daughters, as those images emerged from the projections of different respondents. (From this point of view, the present study used the method of the cultural anthropologist, whereby respondents act as informants, not primarily about themselves, but about one or another datum of social interaction. In this case, the data of social interaction dealt with the social roles of adults in the family.)

After the role-images had been delineated, the investigators turned to possible implications in the data as regards the personalities of the respondents themselves. Thus this investigation broadened in scope as the research progressed and, as will become clearer in following sections, this report deals not only with familial roles but also with the relations between role-image and personality in middle age.

THE SAMPLE

The respondents used in the present study were taken from a larger sample of 240 men and women aged 40 to 70 residing in the metropolitan area of Kansas City. The larger group

constituted a random sample of individuals (not married couples) drawn by area-probability techniques, but stratified by age, sex, and socioeconomic level. The sample had been designed to include successively fewer persons in each five-year age interval from 40-44 to 65-70.

To study possible differences between social class groups, those cases were selected from the original 240 who were most representative, on the one hand, of upper-middle-class patterns of life, and on the other hand, of working-class patterns. All those who fell at the lowest social levels (lower-lower class) were omitted; and all those who, on indices of occupation, education, income, house type, and area of residence, were marginal between upper-middle- and working-classes. Omitted, also, were all non-whites.

These selective procedures yielded 131 cases, distributed as shown in Table 1.

Men

The 32 men of middle-class status are well-to-do business and professional men. There are a few high-salaried accountants and salesmen, executives of large businesses, engineers, lawyers, one judge, and one minister. None are retired, and all but seven consider their present positions to be the top positions they have ever held. With only one exception, all have had at least a high school education, and the large majority have gone to college for two or more years. A few hold higher academic or professional degrees. The origins of these men are varied in terms of social class: some began life in middle-class families, others in lower-class families.

The working-class men, as a group, represent the stable "blue-collar" worker. The majority are industrial workers of varying job rating—assembly-line inspector, oiler, truck driver, machinist, railroad switchman. A few foremen are included, but these are men who have worked their way up through the lower job echelons. Of the 36 in this group, only 3 are retired; and only 9 report that they held better jobs in the past than in the present. The majority of these men finished grade school, a few did not. Some have high school and/or trade school education. (Four attended college, but for less than one year.) All the men in this group are of lower-class origins (thus contrasting with the middle-class men, many of whom have been upwardly mobile).

There are no group differences between middle-class and working-class men as regards marital status, parenthood, ethnic background, or religious affiliation. All 68 have married, and with the exceptions of one who has been widowed and one who has been divorced, all are presently living with their wives. (For 7, this is a second marriage; for one, a third marriage.) All but 8 are fathers, (with families that range

TABLE 1
THE SAMPLE

Social Status	Age 40-54		Age 55-70	
	Men	Women	Men	Women
Middle-class	18	22	14	13
Working-class	21	12	15	16
Total: 131				

in size from one to five children in the middle-class and from one to ten in the working-class) and 25 are grandfathers.

With but one exception, all the men in both social class groups are native-born and are of North European ethnic backgrounds. The great majority grew up in Kansas, Missouri, or neighboring states of the midwest.

Both social class groups show the same range of religious affiliations: the large majority belong to one or another Protestant denomination; six in all are Catholic; one is Jewish; and eight report no church affiliation.

Women

Of the 35 middle-class women, the large majority are married housewives, living with their husbands. (Four have never married; three are widowed, three are divorced.) Their husbands are mostly professional or business men. Of the women living with their husbands, seven hold jobs outside the home, and in three of these cases, the job is one of assisting the husband in his business or professional pursuits. Of the 31 women who have married, six have remained childless; the others have from one to six children. Twelve are presently grandmothers. The women tend to have somewhat less education than men of comparable status, yet there are only nine women in this group who have not gone beyond high school; all the others have had one or more years of college. As with the middle-class men, some of these women were born into lower-class families, others into middle-class.

Of the 28 working-class women, all have married (one is presently divorced, six are widowed, five are married for the second time). All of them married men who are now employed as industrial workers or as blue-collar workers of similar occupational status. The great majority of these women are housewives; only four hold jobs outside the home. All but one have children. Twenty of the 28 are grandmothers.² All are of lower-class origin.

² For both men and women, grandparenthood occurred more frequently before age 55 in working-class than in middle-class.

The two social class groups are similar to each other and to the men in terms of ethnicity and religious affiliation. All 63 women are native-born; all are of North European extraction; all but a few have spent most of their lives in the midwest. The range of religious affiliations is the same as that for men.

THE DATA AND METHODS OF ANALYSIS

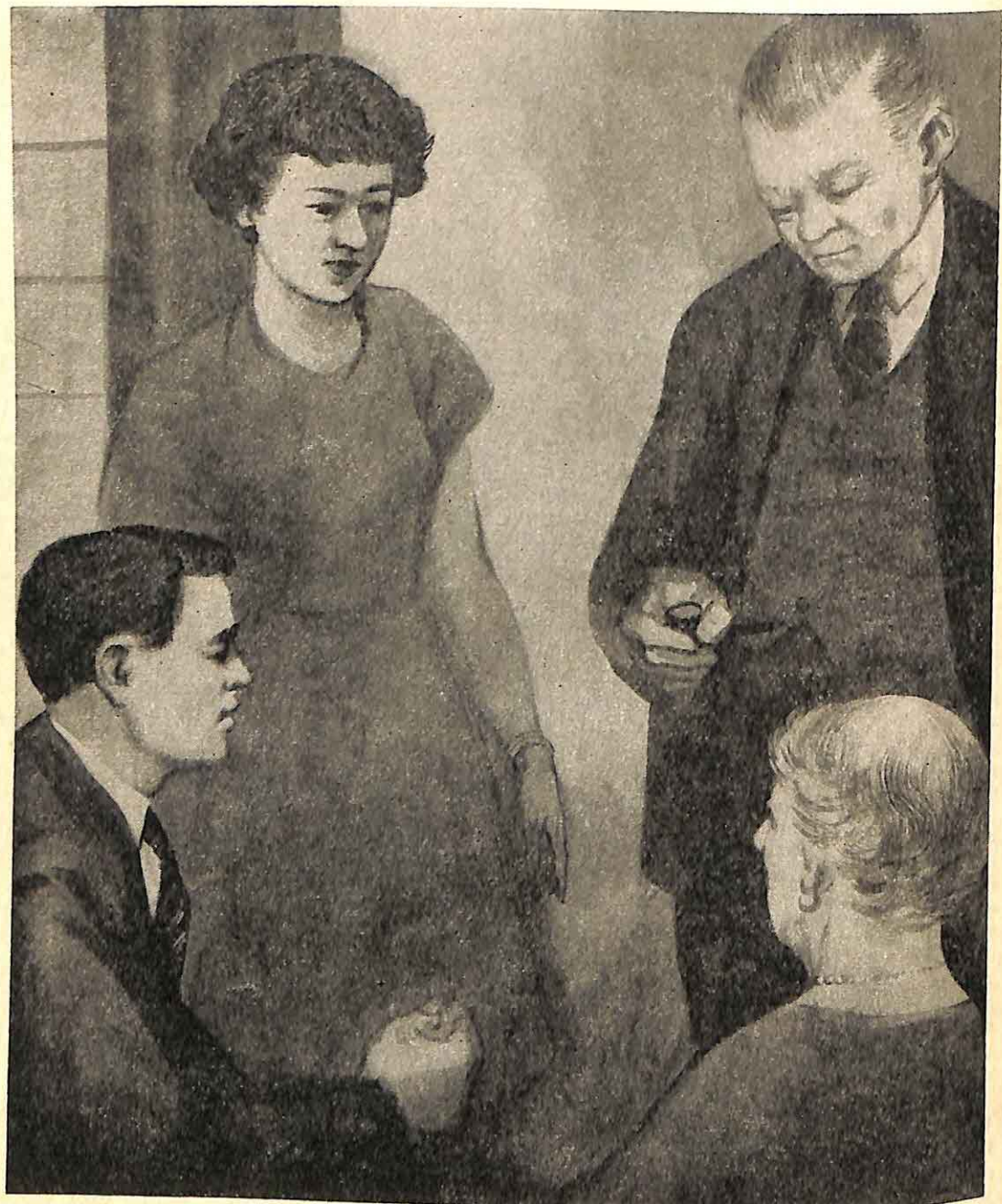
The standard TAT has no single picture appropriate to the purpose of studying age-sex roles in the intergenerational family setting. While it is possible to explore these roles through the standard TAT, the investigators sought a more economical research tool. A specially-drawn picture was therefore used, one designed specifically to evoke the sentiments and preoccupations of middle-aged respondents in relation to family roles.

The Research Instrument

The picture shows four figures, all drawn with a minimum of detail. Observe that at the left is a Young Man seated; then moving clockwise, a Young Woman, standing; an Older Man, standing; an Older Woman, seated. The Young Man is drawn in profile, with the upper torso and right arm shown. He sits opposite the Older Man and his gaze, while vague, is in the direction of the Older Man. The Young Woman, standing near the Older Man, is looking down toward the Older Woman. The Older Man, standing at the right of the picture, is the most prominent in terms of the space occupied in the drawing. His head is slightly inclined, and he holds a pipe. The Older Woman is seated with her back to the viewer. Only the profile of her cheek and jaw can be seen; and while her figure occupies the bottom center of the drawing, it occupies less space than any other figure. While the interaction is ambiguous, the four figures form a group (all four face inward) and together they fill almost the total space of the drawing.

Pretests

Before using the picture in the present research, it was pretested in two different ways. First, a check was made against the possibility that, through some artifact in the way it was drawn, the picture might yield only stereotyped responses. It was tried with men and women of various ages in various informal settings. The responses proved to be greatly varied in tone and content. It was thus concluded that the



Copyright, Committee on Human Development,
University of Chicago

ADULT FAMILY SCENE

picture was successful as a projective technique insofar as it elicited wide individual differences in response.

Second, a check was made against the possibility that the picture presents the respondent with too limited a stimulus to yield data appropriate to the problem. Is the response to this one picture representative of the respondent's covert attitudes regarding age-sex roles—or is the response an artifact of the particular picture and the particular method of administration? More specifically, would the respondent's perception of the Older Woman, for instance, correspond with his perceptions of older women as obtained from standard TAT pictures?

This picture was administered along with cards 2, 6BM, and 7BM from the standard TAT, to 36 middle-aged men and women. Analysis was made, on the one hand, of the respondent's descriptions of a young man, a young woman, an older man, and an older woman, as these descriptions emerged from responses to cards 2, 6BM, and 7BM; and, on the other hand, of the descriptions that emerged from responses to the specially-drawn picture. The two sets of descriptions for each respondent were then compared. Using different methods of judging consistency and of quantifying judgments, two independent analysts made these comparisons and both concluded that there was a generally high degree of consistency. It was concluded, therefore, that this picture was providing data that were consistent, but not synonymous, with data obtainable from the standard TAT; but that, for the purpose of studying adult age-sex roles, the special picture had marked advantages over the standard TAT.

Administration of the Picture

The picture was presented to the respondent at the end of a long interview that covered various aspects of his life history, his views of the life-periods that constitute adulthood, and his attitudes toward aging. The interviews were conducted for the most part in the respondent's own home; occasionally, in his place of business.

Three levels of inquiry were employed in using the picture. The person was asked, first, to tell a story about the picture: a story with a beginning, a middle, and an end. Then the interviewer, moving clockwise around the picture and beginning with the figure of the Young Man, asked the respondent to assign an age and to give a general description of each of the four figures. Again moving clockwise, the respondent was asked finally to describe what he thought each figure in the picture was feeling about the others.

Stimulus Value of the Picture

Almost without exception, all respondents saw the picture as representing a two-generation family. One of the younger figures, most often the YM, was frequently seen as being outside the primary group, usually in the role of suitor or fiancée, or son-in-law or daughter-in-law.

While always structured as a family situation, the stories varied widely in content and in tone. It might be a story of a young man coming to ask for the daughter's hand in marriage and being opposed by the older woman; it might be a mother, father, daughter, son-in-law having a casual conversation before dinner; it might be a young couple asking for financial help from parents; or it might be an older couple coming to visit the younger.

Generally speaking, the two generations were seen as being some 30 to 35 years apart in age. The YM was almost always described as being somewhere in his twenties (the median age ascribed was 25); the YW was always seen as being a few years younger (the median age ascribed was 21). The two older figures were seen most often as being in their late 50's or 60's, with the OW "about the same age" as the OM (the median age ascribed to both the OM and OW was actually 60).³

It is within this over-all setting of the two-generation family, then, that the

³ Contrary to expectation, median ages ascribed to each of the four figures were consistent from one group of respondents to the next. Thus the oldest respondents saw the YM as a 20-year-old as frequently as did the younger respondents. Or, to take another example, the OW was said to be 60 as frequently by women respondents who were themselves 45 as by women respondents who were themselves 65.

role-images of the YM, YW, OM and OW emerged.

General Approach to Role Analysis

Having used the three levels of inquiry, various kinds of data were available for each figure in the picture. The figure was seen first in action categories, as his actions related him to the immediate issue before the family. He was seen, second, in more general or non-situational terms, as he related to a more abstract social or moral environment. He was seen, third, as he related in terms of affect to other significant figures in the family constellation.

The data for each figure can be regarded as a set of related expectations—for our purposes, a set of expectations as to how each category of person (YM, YW, OM, OW) relates to the social environment and to other categories of persons in the family in terms of action and affect, manipulation and feeling. It is this set of expectations which were regarded operationally as the role description. The following assumption was made: granted that the different attributes ascribed by the respondent to the four figures in the picture have their roots in intrapsychic determinants, still the respondent's expectations, based on his experiences with real people, will have a highly determining effect on which aspects of the self he chooses to ascribe to each figure in the picture. In other words, the investigators took the respondent's perceptions as projections, mindful of the fact that what was given was intrapsychically determined, but trusting that the interactional social reality had called out and directed the projection; and that the two (intrapsychic factors and social reality) were syntonetic in the phenomenal field of the respondent.

If this assumption is correct, then respondents did not distribute aspects of the self randomly among the four figures, without regard for the social stimulus value of each figure. It follows that, to the extent there is similarity in the role behaviors of real-life people, there should be a corresponding similarity in the role expectations for such categories of people as they are perceived in the picture.

Illustration of the Method

The first step in the analysis was to see what was the preoccupation around which the respondent had built his story. This preoccupation, as he builds it into his narrative, pervades and colors all aspects of the story; and the role descriptions took on greater meaning once the basic theme, or preoccupation, was understood.

For example, a woman tells the following story:

"I think the boy is going away to the service. He's telling the mother and father. That's his wife with him. The father is pretty downhearted about it. He has a downcast look on his face. His wife doesn't feel too good about it, but she's trying to pacify the older couple. They've just been married. I can't tell how it'll end up. If he has to go overseas to fight, there's always the possibility he won't come back."

General description of the YM: "Sort of a boy who has always been close to his parents. Looks like a nice kind of boy."

General description of the YW: "Looks very sympathetic. She's a real nice girl. She's trying to sympathize with the old people."

General description of the OM: "Looks like a nice home-body. Nice fellow."

General description of the OW: "I can't see enough of her face. I couldn't say any more, because there's no face to go by. Sort of refined, from her stature."

YM's feelings: "He thinks they're all right or he wouldn't have sat down. . . . Well, some boys wouldn't care how their folks felt, but he seems to realize that they're hurt."

YW's feelings: "She thinks her husband is a pretty fine fellow, or she wouldn't be trying to sympathize with his folks."

OM's feelings: "He's pleased with the young

folks. He's interested in what his son's going to do."

OW's feelings: "She's in a bad place—can't see enough of her. I really couldn't say."

The theme which underlies this story is the theme of family dispersion, the "empty nest." The respondent tells us, in effect, that the children are leaving home and that they "won't come back."⁴

Looking first at the description of the Young Man, in the story proper we are told what his action is: to leave home, presumably for some dangerous and rigorous extrafamilial environment. His action has emotional consequences for the parents, consequences which he does not seek to mitigate. His face is set beyond the boundaries of the primary group, and his only action within the group is the rather formal one of making this position clear to his parents. "He's telling the mother and father." It is left for his wife to deal with the emotional consequences of the YM's course.

We are next told about the Young Man that: "He has always been close to his parents." The implication is that an earlier relationship of the son to the parents is now ending. It is of interest that the respondent speaks of the YM's affiliation to the parents only when she discusses him in the general, nonsituational context. When she is asked to consider him in relation to the immediate situation and to the actors in it, the theme of remoteness infuses the portrayal. This point is made clear when

⁴The respondent seems to defend herself against the impact of this crisis by proposing that the son does not leave of his own volition. An impersonal force for which he has no responsibility, the Army, has snatched him away. She probably reveals her true preoccupation, however, when she tells us that the older people feel "hurt," that the younger woman feels some guilt towards the older couple and tries to "pacify" them, and that the young couple have just been married.

the respondent is again asked to put the young man back into the interactive context and to discuss his feelings there. Now he emerges as one who, although essentially detached, still adheres to minimal social forms. The respondent has difficulty in ascribing any but the most qualified feeling to him: "thinks they're all right," and "seems to realize that they're hurt." The YM's reaction to the parents' feelings is a relatively intellectualized one: he "realizes" or recognizes their existence, while it is his wife who "sympathizes" with them. Even this modest affective gesture on the part of the YM is doubted, for the respondent goes out of her way to assure us that the YM is not like other boys who "wouldn't care how their parents felt." In the respondent's whole recital, then, she attributes to the YM only minimal and grudging affiliation to the parents.

We can now make this general statement about the perception of the Young Man's familial role: his basic orientation is to rigorous and compelling nonaffective extrafamilial concerns, and intimacy for him is to be found with peers of the opposite sex. Although not too long ago (in terms of subjective time) the primary, parental group was a major focus of his interest, his present role there is governed by moral directives ("superego" demands) rather than by spontaneous warmth. These directives, coming into conflict with his more compelling extrafamilial interests, result in a posture of grudging punctiliousness, of bare attention to formal, socially defined demands (a "this much and no more" attitude). Himself barely participant in the family—although at the same time, a source of concern to the parents—he leaves to his wife the mediation of the emotional issues between himself and his parents. He is generally governed by outside de-

mands, as though those demands were more congenial to his energies and motivations than are the demands made by the parental group.

Turning to the Young Woman, we are given a quite different image, one which scarcely changes from level to level of the protocol. In the story proper she appears as one whose actions are directed toward the parents (as are the Young Man's). The difference between the Young Man and her lies in the fact that while his action toward the parents—telling them that he is leaving—begins and ends his contact with them, her action—"trying to pacify them"—implies a continuing and multifaceted relationship with the older group. The Young Man only tells them about himself, and at the most can only "recognize" the effects of his announcement on the parents. The Young Woman, while she "doesn't feel too good about it," does not deal with her own reaction but attempts rather to alleviate the parents' grief. The word "pacify" implies maternal behavior, as does the pattern of dealing with the woes of others rather than with her own.

Our interpretation of the Young Woman's strong, maternal concern with intimate human relations is strengthened when we look at the general description given of her. Here, where the respondent's implicit task is to describe the figure in less immediate and less situational terms, the respondent persists in seeing the YW in relation to the current crisis. Again she is seen as a person whose actions are nurturant and consoling, but at this point an element of emotional distance enters into the description. The Young Woman is "trying" to sympathize with the older group. Here is an implication of some barrier against emotional

rapprochement between the old group and the YW, a barrier which she feels impelled to overcome. The description of the YW's feelings give us a clue as to why she attempts to overcome the barrier. It is because of her regard for her husband that she feels a responsibility to his parents. We are told that her primary affiliation is to her husband, and that responsibility to the parents is secondary, stemming from her marital tie. She takes form now as a person who must deal nurturantly with various aspects of the interpersonal universe, although institutional and generational barriers may exist between her and certain others. If formal ties exist between her and other people, she then seeks to enrich the formal ties with empathic bonds.

Taking the role description as a whole, and seeing it against the theme of "empty nest," we see the Young Woman's role as one of emotional liaison, operating in the widening breach between the parental and filial generations. Her husband moves off into what are viewed as distant and threatening events, while she, though drawn after him, bridges the gap and maintains some version of the lost emotional ties between parents and children. Against the background of traumatic family dispersion, her role has a maternal quality: although her primary tie is to her husband, her immediate concern is for those who have been injured by the course of events, and she attempts in maternal fashion to compensate the injured through her nurturance. In sum, her role is complementary to her husband's in that, while he moves off to "do battle," she tarries behind to handle the human consequences of his actions and decisions.

Turning to the figure of the Older Man, we see him first as feeling sad at

the news of his son's leaving. No actions are ascribed to him. He reacts to traumatic situations with feeling, but he is not seen as acting out his feelings or doing anything to alter the situation which made him sad.

At the level of general description, we are told that he is "a nice home-body," positively regarded. His major cathexis or emotional investment is to the family, and it is there that he is gratified. Values relevant to his role are those of comfort and ease in an affiliative setting. (We give "comfort" and "ease" priority over the implication of affiliation, since the description was given in terms of "home" rather than in terms of "family.")

Moving to the feelings ascribed to the Older Man, we notice a shift away from the initial description given of him. Where he was initially saddened, he is now "pleased" with the young people, and he will maintain a meaningful, although somewhat intellectualized interest in his son's future activity. There is still no intimation that he will act to change the course of his son's affairs—the son will "do," and the father will be interested—but a note of equanimity, based on denial, has entered the description. If we examine this shift in the Older Man's feelings in light of the theme of family dissolution, we conclude that this shift represents a concept of defensive adjustment, adjustment to the inevitable and inescapable reality of the Young Man's maturity through defensive denial of strong personal feeling. After some initial depression, the Older Man resigns himself to the fact of the breach and returns to the emotional status quo. While the young group, especially the YM, are no longer reciprocally affiliative, the Older Man's outward feelings toward them remain basically unchanged.

The Older Man's role, then, somewhat like the Young Woman's, is an adjustive one. He buffers the shocks of transition. Accepting the reality of change, he acts to minimize the consequent feelings and to find new bases for intimacy in the new situation. It is of particular interest that the Older Man attempts to maintain the status quo by changing himself and hiding his own feelings. At no point in the protocol does he act to change anything outside himself.

This interpretation of the Older Man's role gains support when we turn to the Older Woman. In the story proper we are told nothing about her; her presence is merely noted. At the general level of description, perception of her is again denied. In effect we are told that she has no emotionally expressive surface ("there's no face to go by"). She is associated only with "refinement," a description which implies that she has no contact with a freely affective, spontaneous environment. (By contrast, the Older Man is the "homebody.") The word "refined" suggests the values of restraint, pride, and possibly a defensive rigidity.

As to the Older Woman's feelings, the respondent in effect tells us that she cannot imagine any feeling states which might pertain, because the OW is "in a bad place." While there is a relative paucity of data about the OW, we nevertheless obtain the strong impression of rigidity and withdrawal in the figure. Viewed against the thematic background, this rigidity and withdrawal takes on meaning as a possible mode of coping with crisis, but a mode which is quite different from the one defined for the Older Man. Faced with the trauma of family breakdown, the Older Man's role is to adjust to the inevitable by minimiz-

ing his own reactions. The adjustive mode ascribed to the Older Woman seems to stress denial of the trauma and its emotional consequences, strict control, and magical defenses against her vulnerability. The Older Man's adjustment, although a defensive one, still is oriented toward a social universe and untroubled contact with others in future situations. The Older Woman's role has a more primitive, egocentric quality, as if the vulnerability of the self were the only concern, and this concern justifies the use of archaic defenses—such as complete denial of a painful situation.

Granted that both the Older Man's and Older Woman's roles may represent possible solutions of the respondent's own problem—her defenses against the problem she has proposed—the investigators' primary interest is with the content of the roles as they emerge from the respondent's fantasy, and to which figure in the picture each role is ascribed. The respondent has presented one role as being appropriate to the father; the other, as appropriate to the mother. It is such concepts of "role-appropriateness" that constitute the data of the present research.

TREATMENT OF THE DATA

Using the method illustrated above, each protocol was analyzed for role descriptions of each of the four stimulus figures. Interpretations were recorded separately for each figure. The protocols were divided according to the sex of the respondent, but were analyzed without knowledge of the respondent's age or social class.

Reliability of Interpretations

The question of reliability of interpretations was dealt with at an early stage

in the research. Using nine protocols selected at random, each of the four figures was rated on a five-point scale for each of twelve personality characteristics: on the extent to which the figure was seen as having strong affect, as being nurturant, succorant, affiliative, deferent, hostile, narcissistic, extra- or intrafamilially oriented, impulsive, autonomous, reality-oriented, and "superego"-oriented (acting on the basis of moral imperatives). Thus, for each protocol, forty-eight separate ratings were made. To keep the ratings as independent as possible, all the OM figures were rated first; then all the OW figures; and so on.

Two judges rated each of the four figures on the nine protocols, and the ratings were then correlated. (One judge was one of the present authors, the second, a graduate student in psychology who had had no previous acquaintance with these data.) The resulting coefficients of correlation were: .81 for the YM figure, .88 for the YW, .83 for the OM, and .88 for the OW.

This procedure proved to be a somewhat oblique test of reliability, since it was later decided not to deal with the role descriptions on the basis of such ratings, but rather to continue to draw summary descriptions of the figure and then to categorize the descriptions according to similarity. At the same time, this test of reliability was a relatively stringent one for this type of data.

In the light of the high correlations, it was concluded that the investigators' interpretations were not idiosyncratic, and that an independent judge would see in the data essentially the same patterns that we saw.

Quantifying the Data

Once all the protocols had been an-

alyzed, then decoded, the data for each figure were treated separately. The procedures with reference to the data on the Older Man will be described here since these were the first to be dealt with, and since findings regarding the OM influenced in some ways the organization of the other sets of data.

All the descriptions were grouped on the basis of similarity into mutually exclusive categories, attempting always to judge similarity in terms of the most salient features ascribed to the OM by the respondent. This was a lengthy process since the attempt was to establish categories that would produce the least distortion of the original data. At the same time, having become aware that there were age differences in the perceptions of the OM, the investigators attempted to structure the categories in such fashion as to highlight the age differences.

Six major categories were finally delineated and arranged along a continuum termed "dominance-submission." At one end were those categories in which, whatever other characteristics were ascribed the OM, he has always been seen as a dominant figure in the family. At the other end were those categories in which he was seen as a passive or submissive figure. Dominance or submission was judged in terms of the OM's impact on the situation; the extent to which others deferred to him; the extent to which resolutions of family issues depended upon his wishes, his judgments, or his decisions.

Dealing next with one after another subsample of respondents (middle-class men, middle-class women, working-class men, working-class women), frequency distributions were made in which the role descriptions were plotted by cate-

gory and by age of respondent. The distributions were then tested for statistical significance by applying the chi-square method.

For each subsample of respondents there was a shift with age in the perceptions of the OM. The role descriptions given by younger respondents (aged 40 to 54) fell more often in those categories in which the OM is described as dominant; the descriptions given by older respondents (aged 55 to 70) fell most often in those categories in which the OM is passive. The number of cases in each subsample was too small to establish reliable chi-square values, but the trend was present in every group. Cases were then combined into larger groupings—all male respondents, all female, all middle-class, all working-class. The age trends were now even more pronounced (P values were between .05 and .01). Finally, when all respondents were grouped together, the age-shift was unquestionably reliable (P was .001).

At this point, while age of respondent consistently produced variation in the data, it was not clear which of the original variables—age, sex, or social status—was the most important in producing the overall variation. A further step was therefore taken. Ratings on dominance-submission were assigned to each category of role description; these ratings were then submitted to an analysis of variance. It was found that of all the variables—age, sex, social status, and the interactions thereof—only age was significant (P was beyond .001).⁵

⁵ The data on the Older Woman were also subjected to analysis of variance, with the same result emerging. Data on the YM and YW could not be treated in the same fashion, however, since the categories in these two sets of data could not justifiably be rated along a single continuum.

Presentation of Findings

In the sections to follow, the data for each of the four figures are reported in turn. For each figure there is first a general description of the role-image as it emerged from the total array of protocols; then a discussion of age, sex, and social-class differences among respondents in their perceptions of the figure. Finally there is a discussion of the implications of these differences in regard to the personalities of respondents.

In moving from the role-images to the implications for personality, the investigators are aware that they move to a different area of interpretation. An example may aid in clarifying this point: the Young Man's role is perceived differently by 40-year-olds than by 60-year-olds. This variation may, on the one hand, reflect different social realities for 40- and 60-year-olds (i.e., a Young Man may in reality behave differently in interaction with a 40-year-old than in interaction with a 60-year-old). On the other hand, this variation is likely also to reflect differences in personalities of the respondents who tell the stories. (The raw data, it will be recalled, are fantasy data and, as stated earlier, it must be assumed that the individual's personality partly determines the role images.) Having described the age differences in the images of the YM, the investigators turned to the second area of interpretation and asked what these differences imply regarding the personalities of 40-year-olds and the personalities of 60-year-olds.

The statements dealing with personality are offered as speculative rather than conclusive. At the same time, the investigators believe that exploring the data in this second fashion has produced hypotheses regarding adult personality

that merit further research and verification.

THE YOUNG MAN

Of the four figures in the picture, the Young Man was ascribed the least significant familial role. While all respondents saw the picture as a family setting, and while the other three figures could be comfortably described within this setting, this was not the case with the YM. Almost half our respondents saw him as being oriented primarily toward the extrafamilial world and only secondarily toward the family.

Within the family the YM is a rather remote, detached figure; his membership in the group depends more upon formal than upon affective ties. In both the family and the outside world, social forms and institutions, rather than qualities intrinsic to himself, define the YM's role and guide his actions. The impression is that the YM exists in a set of institutional matrices, and it is the workings of such matrices that determine what happens to him. For instance, the YM is often seen within the family in the role of suitor of the YW. He makes his request and establishes the problem; but it is then a problem that is out of his hands and depends for its resolution upon the interaction between the OM and OW.

The YM's only source of strength is that he represents the more or less legitimate demands of the extrafamilial social environment, and accordingly his claims must be treated with some respect and consideration.⁶ His demands are po-

⁶ In 92 stories told about the picture, the YM was seen as posing a threat to the primary group—he is asking for the daughter in marriage, or he is about to leave home, or he and his wife are thinking of moving to another city. In most of these stories, the YM's demands are being opposed by the Older Woman (who is de-

tentially disruptive to the unity of the primary group, but they symbolize ambition, independence, and maturity, and cannot easily be denied.

A frequent component of the YM's role is controlled aggression. If he is seen as potentially self-assertive or intrusive within the family, it is implied that, by exercising self-control, he must protect the parents from the possibly destructive consequences of his impulses. Or, if he does act assertively, he must nullify his action by propitiating the older people. Aggression in the YM can be freely expressed only when it is directed outside the family and when it takes the form of ambition or achievement drive. The YM's aggression is never antisocial.

It is not only aggression that is neutralized within the family; there are also few spontaneous affiliative feelings ascribed to the YM. His emotional expressivity within the family is generally restricted for all categories of affect. This interpretation is supported by the fact that only 29% of respondents (in Categories 3 and 5 of Table 2) describe an openly affective familial role for the YM, while 71% see him primarily in terms of achievement strivings, social conformity, and impulse control (Categories 1, 2, and 4 of Table 2).

A salient feature of the YM's role, then, is that he responds most to promptings from the outer world, and least to inner prompting or impulsivity. The implication is that for the YM it is too dangerous to act directly from impulse; he must find guidance in the safe structure of environmental rules and directives.

scribed in these records as giving priority to the values of intimacy and family solidarity) and is being abetted, at least tacitly, by the Older Man (who is seen as being more in touch with outer-world demands than is the OW).

Categories of Role Descriptions

In examining the role descriptions of the YM, it appeared that there was no single continuum nor single variable by which the data could be ordered. There were at least two major dimensions of the role that seemed to preoccupy respondents and that varied independently. One related to the YM's "psychological location," whether he was seen as being oriented primarily toward the outside world or toward the family. The second related to the vitality ascribed to the YM, the extent to which he was perceived as acting on the basis of impulsivity, drive, inner motivation.

Five broad categories provided the best fit for the data. They are described here in the order in which they appear in Table 2, where they have been arranged to highlight the age differences between respondents.

1. Outer world achievement. Here are grouped all those role descriptions which ascribe to the YM qualities of aggressiveness or assertion directed toward the extrafamilial world, the commercial-technical environment. (In this category, a typical description of the YM is "ambitious.") The YM is primarily oriented toward the outer world and it is there that he is seen as expressive and achieving.

Within the family, the YM is affectively remote, bland, or deferent, and he does not emerge vividly in any categories of affect or action. If self-assertive needs within the family are recognized at all, they are seen as qualified or attenuated. If affiliative needs are recognized, there is a lack of strength or spontaneity about them. The YM reacts to the family (especially to the parents) in moral and rational terms, rather than in affective terms. He "respects them" and "values their advice;" he rarely loves them. Any interpersonal bond is, by implication, formalized.

2. Family: controlled. Here the YM is primarily oriented toward the family and has strong needs for assertion within the intimate family group; needs which must be controlled and limited. The control resides either in parental authority, or in the YM's own internalized controls. Thus the YM is seen as checking his aggressive energies and thereby protecting the

TABLE 2
ROLE DESCRIPTIONS OF THE YOUNG MAN

Category:		Age 40-54		Age 55-70	
		Men	Women	Men	Women
1. Outer-world achievement	Middle-class	8	7	1	0
	Working-class	7	1	2	0
	Total	23*		3	
2. Family: Controlled	Middle-class	3	8	0	6
	Working-class	7	4	3	3
	Total	22		12	
3. Family: Affiliative	Middle-class	1	2	5	2
	Working-class	2	2	1	1
	Total	7		9	
4. Outer-world conforming	Middle-class	1	3	2	3
	Working-class	2	4	5	9
	Total	10		19	
5. Family: Assertive	Middle-class	2	2	3	2
	Working-class	3	1	4	3
	Total	8		12	

* The chi-square test was applied to these category totals. The probability that the distribution occurred by chance is less than .001.

older group from its consequences. If he acts assertively, he must undo his action by reverting to a supradeferent or abasive stance.

3. *Family: affiliative.* The YM is seen here primarily in terms of a potent and spontaneous cathexis to the older group, with strong needs for intrafamilial affiliation. The affiliative needs are usually tempered with succorance needs. (The relative infrequency with which this category occurs is in contrast to the data on the YW.)

4. *Outer-world conforming.* This category is similar to No. 1 in that the YM is primarily oriented toward the extrafamilial world, but here he is seen, not in terms of an energetic, intrinsic drive toward achievement, but rather in terms of qualities *relevant* to achievement and success. He is "intelligent," "clean-cut," "well-groomed," rather than "ambitious" or "aggressive." He is trained for reliable performance, rather than driven to achievement.

Within the family, as in category No. 1, he is affectively bland, restrained, remote, and relates to family members in rational and moral terms rather than with spontaneous affect.

5. *Family: assertive.* Central here is the perception of the YM as an assertive peer in the family. He defines his own needs, goals, pur-

poses, and without propitiation, he works to further them within the family. He is expressive, and chooses freely the objects of affection and dislike among the older group.

Age Differences

As will be seen from Table 2, Categories 1 and 2 occur predominantly among younger respondents (aged 40 to 54); Categories 4 and 5 occur predominantly among older respondents (aged 55 to 70). The 40-year-old tends to see the YM as energetic and self-propelling. His aggressive, intrusive energies are funnelled towards the extrafamilial environment, where he is granted the right of self-expression and achievement; but these energies are successfully checked and controlled within the family setting.

For the 60-year-old respondent, the outer world is still the important area for the YM (Category 4), but it is no

longer an area which passively awaits the thrust of the YM's dynamic energies. Instead, the outer world is one into which the YM fits neatly by virtue of possessing those qualities which match the demands of a complex environment. The older respondent is still concerned with the YM's aggressive energies within the family (Categories 2 and 5), but now there is less faith that these energies can be controlled or profitably channelled into extrafamilial achievement.

The general age shift, then, is from seeing the YM as energetic in the outer world to seeing him as intrusive within the family; and from seeing him as aggressive and achieving in the outer world to seeing him as passive and conforming in the outer world.

If it is true that the stimulus figure of the YM symbolizes relatedness to the outer world, then the changing ways in which the YM is seen to relate to the outer world imply personality differences in respondents at the two age levels. For younger respondents, the individual possesses energy congruent to the opportunities within the outer world. A certain predatory quality is present in the individual, and the environment is seen as rewarding boldness and risk taking. There is a simple equivalence: one gets from the world what one puts into it.

Older respondents seem more cautious. In viewing the YM as prepared for, rather than motivated towards, achievement, there is the implication that the outer world is large and complex. One does not reform the world in line with one's wishes; one rather conforms to it. There may be some fear of the outer world among older people—a feeling that the environment is dangerous and that it is best approached with respectful and cautious gestures. Furthermore, to the

extent that the YM is himself a symbol of the entering wedge of the outer world within the family, the aggression that older respondents assign to him perhaps represents their fear of a hostile and encroaching outer world—a "young man's world."

Sex differences

The differences between men's and women's views of the YM stem from their concerns with control versus impulsivity. Categories 1, 3, and 5 in Table 2 represent relatively free affective expressivity for the YM, either in the family or in the outer world, while Categories 2 and 4 represent measures of conformity and restraint. Men's perceptions of the YM are clustered in the expressive categories, women's in the control categories. (The difference between the sexes in this regard is stable beyond the .005 level.)

Within the family setting (Categories 2 and 5), men give more stress to the YM's unhampered interpersonal assertiveness (Category 5), while women place greater stress on the note of control (Category 2).

Similarly, of Categories 1 and 4, men use Category 1 where the YM is achieving and self-motivated in the outer world; women use Category 4 where the YM is only functionally oriented to the outer world. Thus men see the YM in terms of intrusive, aggressive drives, while women see him in terms of plasticity and sensitivity to outer pressures. For men, the YM pushes energetically into an unstructured environment. For women, the YM is embedded within a complex, highly-articulated environment—he fits neatly and without strain into an established order that has already molded him to its needs. It is as if male respondents propose that the YM is a bull, while

women respondents say that he cannot be a bull because the world is a china shop.⁷

In general, then, men tend to accent the energy that the YM brings to role performance. Women recognize the energy, but obliquely, through the stress they place upon controlling it.

Social-Class Differences

While examination of the data in Table 2 shows certain differences between all four subsamples of respondents in the way their images of the YM are distributed, the most striking difference relates to working-class women. Of 31 responses from working-class women, only one falls in Category 1, while 13 fall in Category 4. Thus the working-class women seem unable to conceive of the YM as an aggressive achiever, internally-motivated; they stress, instead, the outer demands and influences which bear upon him. For these women, the YM is "bleached-out," affectively remote, with little strength or drive in either the intra-familial or the extrafamilial worlds; he is related to others in terms of morality and rationality rather than through patterns of impulsivity or intimacy.

THE YOUNG WOMAN

The Young Woman, by contrast with the YM, is not only comfortably accounted for within the setting of the family, she seems to live only within that setting. Respondents perceive no roles for her other than familial roles.

She is seldom a sharply-etched figure. "Nice" or "pleasant" are words frequently used to describe her, and she is

⁷Female sexual anxieties may be involved here—the need to limit and confine the intrusive energies of the male, and to make of him a "nice" person.

most often seen as "waiting" for a decision to be made by the others. Few respondents attribute to her great vitality or autonomy or determination (she stands in contrast to the OW in these respects). At the same time, the YW occupies a central role in the family. While she is not a vigorous figure who initiates or resolves family issues, she is the figure about whom issues are initiated and resolved. Typical stories are those in which the YM wants to marry her and she waits for her parents' (especially her mother's) decision; or she is already married and about to have a child and the issue is now what her parents will think of the news; or the older couple is giving advice to the younger and she is the one who sees both sides of the issue.

It is through the YW that the YM is usually related to the older figures; the YW who provides the tie between younger and older generations.

The Young Man, it will be recalled, is seen as bridging the familial and the commercial worlds; in somewhat similar fashion the Young Woman bridges two systems of affiliation: affiliation to opposite-sex peer, and affiliation to parents. Respondents seem preoccupied with the issue of the YW's affiliation. To whom does she or should she belong? How should she divide herself between the roles of woman and child? What is the nature of her affiliative bonds—are they spontaneous and affective, or formal and deferent?

As regards respondents' personalities, we have suggested that the YM symbolizes the problems posed by intrusive masculine energies. The YW seems to symbolize the problems related to tenderness, intimacy, and sexuality—how much are these to be cathected to heterosexual objects and how much to pregenital objects? For

older respondents, a further problem is, given that the YW represents a libidinal fount, how much love can they expect to receive? Or what substitutes for this love can they legitimately claim?

Categories of Role Descriptions

The roles descriptions of the Young Woman can be subsumed under six broad categories, each representing a different pattern of the YW's affiliation with the other figures in the picture.

1. *Freedom from parents.* Here the YW is strongly affiliated with the YM (her fiance or husband). She strives successfully against the OW for the right to take on adult sex roles (wife, homemaker, mother); or, already filling such roles, she fights off the encroachments of the OW (usually defined as her mother-in-law).

2. *Affiliated to parents.* In this category the YW is seen as related affectively only to the parents. Childlike, she has strong and spontaneous needs for affiliation only with them. Even in those stories in which the YM is seen as her husband, her relationship with him is described in formal terms only, and her primary ties remain to the parents.

3. *Complementary affiliations.* Here the YW is affectively related to both the YM and the parents. She has her feet in both camps—she is a good wife and mother, she loves her husband, she loves her parents, and there is no incompatibility among these loyalties. Her affiliations are complementary and nonconflicting.

4. *Conflicted.* In these stories, the YW has strong needs for autonomy, away from the parents and toward the YM; and at the same time, strong needs for deference and succorance towards the parents. She is seen in what might be called an adolescent role, as one who pulls forcibly away from the parents toward mature roles and heterosexual affiliation, while at the same time remaining dependent upon them. She is torn between childish and mature role choices.

5. *Deference to parents.* Here the YW relates to parents in terms of succorance and deference, but the relationship is formal, rational, and without warmth. She remains affectively remote from parents and her ties with them are based upon "superego" demands rather than upon libidinal demands.

6. *Controlled by parents.* Here the YW appears in much the same terms as in Category 5—deference or succorance toward parents, without affective affiliation—except that here there is an

added element of being controlled by the parents. These stories deal always with the situation in which the YW is considering marriage (thus at least a formal need for heterosexual affiliation away from the family is recognized) and is deferring to her parents' decision. Not bound to her parents by love, the YW nevertheless passively yields to them the disposition of her affairs.

The role descriptions of the YW are distributed as shown in Table 3.

Age Differences

As seen in Table 3, Categories 1 and 2 represent the images of the YW held predominantly by younger respondents (age 40-54), Categories 3 and 4 occur about equally frequently in both age groups, and Categories 5 and 6 are those of predominantly older respondents (aged 55-70).

Younger respondents see the YW in polar ways. She is either pulling assertively away from the family, intent on becoming a woman and a mother in her own right, or she is freely and spontaneously relating to the parents alone, without cathexis to the YM and without desires for mature womanly achievement. Younger respondents, as a group, are undecided about the nature of the YW's affiliation, and the claims that parents can make upon it. There is tension over the issue of allegiance to an extra-familial (in this case, the YM) or to an intrafamilial identity.

On another level of interpretation, that pertaining to the personalities of respondents themselves, we may infer some conflict in younger respondents between independent and dependent modes of cathexis; between sexually mature and immature tendencies.

For older respondents, the role descriptions have in common the YW's lack of affective ties with parents, but her deference towards parental authority.

Perhaps the contrast between Categories 1 and 6 reveals most sharply the nature of the age shift in perceptions of the YW. Category 1 presents the YW as strongly oriented toward heterosexual affiliation and as desirous of becoming a woman in her own right. Parental controls, if they exist, are to be flouted rather than conformed to. Needs for peer affiliation and mature status are primary. Category 6, on the other hand, shows the YW as being aware of these needs, but now it is parental controls which have primacy. Rather than acting on the basis of her own motivations and her own drive toward maturity, she acts on the basis of parental needs and her own introjected controls.

Older respondents seem to have abandoned the expectation that the YW will have any strong feelings for parents. They stress the submissive rather than the libidinal qualities in the parent-child relationship (to the extent that the YW is affiliative toward the older generation, the affiliation is charged with succorance needs rather than with spontaneous warmth). Instead of love, parents will have respect. To older respondents as a group, the figure of the YW no longer presents the conflict seen by younger respondents—the conflict between heterosexual impulses that have a centrifugal effect upon the primary group and childish impulses that have a centripetal effect. Now the YW's centrifugal tendencies are

TABLE 3
ROLE DESCRIPTIONS OF THE YOUNG WOMAN

Category:		Age 40-54		Age 55-70	
		Men	Women	Men	Women
1. Freedom from parents	Middle-class	5	7	1	3
	Working-class	6	2	1	1
	Total	20*		6	
2. Affiliated to parents	Middle-class	4	3	0	1
	Working-class	3	3	1	2
	Total	13		4	
3. Complementary affiliations	Middle-class	0	3	4	1
	Working-class	4	3	2	2
	Total	10		9	
4. Conflicted	Middle-class	5	5	5	2
	Working-class	5	2	4	3
	Total	17		14	
5. Deference to parents	Middle-class	1	1	3	1
	Working-class	2	1	3	3
	Total	5		10	
6. Controlled by parents	Middle-class	1	3	1	5
	Working-class	1	1	4	5
	Total	6		15	

* The chi-square test was applied to these category totals. The probability that the distribution occurred by chance is between .005 and .001.

seen as controlled, and the YW herself is seen as taking responsibility for maintaining some vestige of the parent-child relationship.

Turning again to questions of personality change with age, it has been suggested that the YW symbolizes the issues of intimacy and sexuality. With increased age of respondent, there is increased emphasis upon the YW's dependence on parents for nurturance and control. For 60-year-olds, then, the conflict between heterosexual and pregenital modes of cathexis appears to have been resolved in favor of the latter. Heterosexual concerns are no longer seen as compelling, and are not viewed as legitimate reasons for breaking away from the home. Perhaps a further implication is that, for the older respondent, the world seems to be shrinking toward the confines of the controlling and nurturing family. Older respondents in fantasy no longer strain away from the family; instead, they seem to identify with the controls that keep one within it.

The older respondents' perceptions of the Young Man are also relevant here. The YM is seen as relating to the outer world as a conformist, as one who adapts himself to the demands of the extra-familial environment. Similarly, the Young Woman relates as a conformist to the parents, as one who lives in a world of strictures rather than in a world of self-initiated action. (This finding, that with increasing age of respondent there is increasing conformity ascribed to the young, is statistically reliable beyond the .001 level.) The similarity in these views suggests that, to the 60-year-old, the individual is no longer a forceful manipulator of the object world, but is instead a rather passive object manipulated by the environment. The individual can de-

termine only partly the outcomes of such manipulations, and then only by adopting attitudes of accommodation and conformity.

Sex differences

Men and women are in general agreement in their perceptions of the YW. Table 3 shows that men more often than women see the YW as conflicted over the question of where her primary affiliations lie (Category 4), and that women more often than men see the YW as being controlled by parents (Category 6); but the overall similarities between the sexes are perhaps more striking than the differences.

At the same time there are certain sex differences in perception of the YW that are not shown in Table 3. Younger men sometimes describe the YW as already established beyond the family confines; as a wife and mother in her own right, and as being affectively remote from parents (these stories form a subcategory of Category 1). This view of the YW never occurs among women respondents. Instead, throughout these data, women always demand that the YW's movement away from the family be accompanied by some compensatory gesture of affiliation or deference toward parents. Women cannot be neutral or unconcerned either about the YW's identifications nor about the more general issue of setting children free. (Something of the same tendency occurs around the figure of the Young Man, although to lesser extent.) Men are somewhat more willing to admit that the YW can be equally a wife and daughter; women tend to say that she is first a daughter and second a wife.

The greater concern among women over controlling the YW (especially true of older women) is additional evidence

for the thesis that women are more involved emotionally with the YW than are men and, by implication, that women are more concerned with the issues of intimacy and sexuality than are men.⁸

Social-Class Differences

Examination of Table 3 shows great similarity between middle- and working-class respondents in their perceptions of the YW. In only one segment of the data is there a significant difference: middle-class women, as compared with working-class, more frequently see the YW in terms of Category 1—as vigorous, autonomous, sometimes rebellious, and successful in breaking her ties to parents.

It will be recalled that these women also ascribe energy and drive to the Young Man more frequently than did working-class women. Thus, while equally concerned with the issue of controlling the young, middle-class mothers, by ascribing greater vitality to both the YM and YW are implying that the relationship between younger and older generations is a complex problem; and that young people, by being something of a match for parents, are not easily controlled.

THE OLDER MAN

The Older Man is sometimes an authoritative, but more often a submissive figure in the family. Those respondents who see him as dominant tend also to propose some hesitations in the exercise of that role. He may act to limit the young people in some way, but if so, he puts restrictions not on their freedom of action but on their succorant demands

toward him. If he is approached for advice, he may tell children in effect to "grow up" or "use your own judgment." In short, the OM is most authoritative when he is pointing out to the others that he is not an authority.

In some instances, the OM is seen as standing between an autonomy-seeking YM and an overnurturant OW who would spoil the YM if the OM did not intervene. This issue of the YM's untrammelled achievement is one of the few about which the OM can act assertively. In most cases, however, he is overwhelmed by the complexity of an emotional situation, withdraws, and turns it over to the OW. At other times, he is isolated altogether.

On the other hand, somewhat like the YM, the OM is not altogether circumscribed within the family setting. It is recognized that he has preoccupations in other areas. While he is seldom described as achieving or successful in the extra-familial world (his is a custodial rather than an achieving role), still the implication is clear that to gauge his impact in the family setting is not to gauge his full worth as a person.

In comparison with the Older Woman, the OM has a certain blandness of affective life. He tends generally to be passive and resigned to whatever happens. The affect that he most consistently expresses is that of uncritical affiliation: "He likes them all and wants to do the right thing for everybody." The OM has, in general, less impulse life than the OW and is less often portrayed as being in conflict situations.

The essence of the OM's role is perhaps this: he can wrestle with the interpersonal environment when such activity has the purpose of gratifying others. If he stands opposed to others, then he

⁸ There is much more YW-OW conflict described in the stories than YM-OM conflict; and while there are few instances of YM-OM conflict over the YW, there are many instances of YM-OW conflict over the YW.

must alter himself to conform to their needs, or he must withdraw to a world of inner contemplation where the actions of others do not impinge upon him.

The figure of the OM seems to symbolize for respondents the ego qualities of the personality: the rational rather than the impulsive approach to problems, concern over the needs of others, reconciliation between opposing interests, cerebral competence.

Categories of Role Descriptions

The descriptions of the OM fall into six major categories that can be ordered along a continuum from dominance to submission.

1. *Altruistic authority.* In this category the OM is seen in a position of authority in the family and he uses his authority to benefit the young people or the family as a whole. He is the benevolent monarch, the nurturant wise man, whose actions are altruistically motivated and lead only to benevolent outcomes. He operates effortlessly and easily in this role.

2. *Assertive, but guilty.* These descriptions are those in which the OM attempts to further his own ends, but is restrained by inner reluctances, doubts, or guilt. He occupies a position of strength and asserts himself within the family; and while he is not opposed by others, he nevertheless cannot easily and singlemindedly press for his announced goals. There is always some quality of inner doubt about the justice of his claims. He is conflicted, unsure, the insecure autocrat. "He thinks it's about time those kids left home and earned their own living—he hates to tell them, though."

3. *Formal authority.* Here the father is the authority, but by default. His authority is challenged as the story progresses; or other individuals take action to decide outcomes while he acts only to approve those outcomes. He is described here not so much in terms of service to others (as in Category 1), but in terms of pliability to the wishes of others. He merely approves decisions which have already been thrashed out among more active figures.

4. *Surrendered authority.* It is indicated here that the OM could be the authority if he desired—he possesses the requisite qualities—but he refuses and/or abandons the role. In some instances, he is initially described as dominant, but as the story unfolds he is relegated to a

more submissive position. In other instances, he is ascribed the qualities associated with leadership, but these qualities are split away from action—they have no impact on the events of the story, they do not impinge on outcomes, they find no overt behavioral expression. He is inwardly "tough,"—but overtly passive; or he has "executive qualities,"—but leaves the decision up to his wife. In no instance is there an intrusion of the OM as a dominant force on the family scene.

5. *Passive, affiliative.* Here the OM is described in terms of what might be called maternal qualities. He is unflaggingly and uncritically affiliative toward the others. He "loves everybody." He accepts, resignedly, outcomes which he may not approve. He is dominated by his wife, but seems to feel no discomfort or resentment in the situation. In stories where the OW is opposing some action proposed by the young people, such as marriage, the OM's attitude is one of affiliating with both sides—of saying affectionately to the OW, "Why fight the inevitable?"

6. *Passive, cerebral.* Grouped here are those descriptions which present the OM as passive and withdrawn. He lacks any announced affiliative attachments to others. The issue of authority does not even arise. His wife rules the family, and he remains remote, both in terms of action and affect, from the family drama. As this drama swirls about him, he "thinks." (The content of his thought, or its relevance to the situation, is rarely specified.) The OM controls events from behind the forehead, as it were, and takes a certain satisfaction in the freedom this provides him. As one male respondent put it, "He's made up his mind about the thing. He's waiting for the old woman to tell them what to do."

Age Differences

As shown in Table 4, there is a consistent shift, with increasing age of respondent, from seeing the OM in situations of power in the family toward seeing him as passive and submissive. (This age shift is statistically significant beyond the .005 level.)

The stimulus figure of the OM confronts the majority of younger respondents with the issue of familial authority. (If category 4 is included, then approximately 75% of all younger respondents see the OM either as an authority figure

TABLE 4
ROLE DESCRIPTIONS OF THE OLDER MAN

Category:		Age 40-54		Age 55-70	
		Men	Women	Men	Women
1. Altruistic authority	Middle-class	5	9	2	1
	Working-class	3	3	1	3
	Total	20*		7	
2. Assertive, but guilty	Middle-class	2	0	2	2
	Working-class	6	1	1	1
	Total	9		6	
3. Formal authority	Middle-class	5	5	0	0
	Working-class	2	3	1	0
	Total	15		1	
4. Surrendered authority	Middle-class	2	3	4	3
	Working-class	6	1	3	2
	Total	12		12	
5. Passive, affiliative	Middle-class	0	4	2	6
	Working-class	3	4	3	8
	Total	11		19	
6. Passive, cerebral	Middle-class	4	1	4	1
	Working-class	1	0	6	2
	Total	6		13	

* The chi-square test was applied to these category totals. The probability that the distribution occurred by chance is less than .001.

or as one who possesses the potential for authority.)

Each of the first four categories in Table 4 represents a different resolution of the issue of male dominance. The first two represent active resolutions. If the issue is met head on—if, that is, the OM defines self-gratifying goals and uses his position of authority to achieve them (Category 2)—then ambivalence and guilt are the necessary results. If, on the other hand, the OM uses his authority nurturantly for the benefit of others, he can act easily and comfortably in his position (Category 1). The more passive solutions involve either the OM's sanctioning of the wishes of others and attempting no intervention in the family scene (Cate-

gory 3), or the more outright abandonment of the authoritative status altogether (Category 4).

For our 40-54 group, the issue being dealt with around the role of the OM is not only that of male dominance, however, but also that of male aggression within the family. (The latter was also one of the primary issues in the role descriptions of the YM.) The problem seems to be how the OM can be an authority without being arbitrarily, and perhaps harmfully, self-assertive. How can the cultural demand—that the father is head of the family—be met without exposing the family to male aggression? As already suggested, the solution seems to involve the stressing of the moral func-

tion of authority: the OM must be either an active force for good; or, by "letting things happen," he passively cooperates with the others in arriving at positive outcomes.

For older respondents, the OM no longer presents the issue of masculine authority. The stories are now those in which the OM has no impact upon family events, and he presents only one or another image of passivity. (Of all respondents aged 55-70, 55% are found in the last two categories. If Category 4 is included—the OM abandoning or surrendering authority—then 80% of all older respondents see the OM among the categories of submission and denial of authority.)

The age shift in the image of the OM from dominance to submission is elaborated in several ways. The forty-year-old respondent sees the OM as being in doubt about his own assertive tendencies; the 60-year-old sees him as being the passive object of others' assertion. In the 40's, the OM is seen as attempting to control events. In the 60's, he only attempts to control and order the cognitive environment, the symbolic traces of objects and events. In the 40's, it is proposed that the OM is aware of the pressures from an impulsive and wilful woman, but that he can allow the OW full expression and still wisely control the course of events. In the 60's, it is proposed that impulse, in the form of the OW, is left in charge of the field, that the OM's wisdom can only control events behind the forehead. The OM has moved from a stance of intrafamilial autonomy to "intracranial" autonomy.

As regards the implications for personality differences in respondents, it has been said that the figure of the OM symbolizes ego qualities of the personality.

With increased age of respondent, the ego, as personified by the OM, seems to contract. On the one hand, it is no longer in contact with impulse life, controlling and channeling it (the OM no longer controls the OW's impulsivity). On the other hand, the ego is no longer in a position of mastery relative to the outer world (the OM is not successful in the extrafamilial world). Ego functions are turned inward, as it were, and while rational thought processes are still important in the personality, thought is no longer relevant to action.

Age-Sex Differences

While age differences in the perception of the OM along the dominance-submission axis are consistent, there are, at the same time, important differences between the sexes.

Women. Among younger respondents, it is primarily the women who see the OM in the altruistic, nurturant role (Category 1), while it is the men who tend to see him in the role of insecure autocrat (Category 2).

Women see the OM as a kind of benign ego figure who exercises his authority benevolently. He allows the children to go off and get married, for instance, and he lets the OW rave about their leaving, knowing all the time that this is what reality is: children grow up, and the OW's feelings about it, no matter how stridently expressed, cannot really change anything nor hurt anyone. In allowing the OW to rant, the OM nevertheless deftly controls her, and he prevents her feelings from dominating the situation. (It is as if the women respondents who tell these stories are saying of themselves that they rely on their husbands to let them be emotional and expressive; that so long as their husbands

are present to control events they, the women, cannot hurt themselves or others.)

Among the older respondents, it is predominantly the women who put stress on the OM's affective qualities, especially his "sweetness" (Category 5). There is a nurturant, almost maternal quality in the OM's passivity.

The OM appears to both younger and older women, then, as one who is identified with the needs of younger individuals. The difference between younger and older women respondents lies in the fact that, for the former, the OM is seen as enforcing his nurturant views on events, guaranteeing benevolent outcomes to the young. For the latter, the OM no longer acts effectively from his nurturant posture. The OM's role does not now have the same quality as before of a barrier interposing between an incensed OW and the vulnerable young. The OW is let loose in the situation, and the OM is no longer capable of checking her. (At the same time, it is suggested that really the young are no longer vulnerable and that they somehow operate beyond the OW's reach.)

The age shift for women occurs, then, in terms of the OM's decreasing effectiveness in implementing his nurturant attitudes and in making outcomes conform to them.

With regard to personalities of respondents, women, in responding to the figure of the OM, seem to be preoccupied with concepts of altruism, nurturance, and generativity. In early middle age (40-54), they propose that these qualities are the important ones in the disposition of affairs. Even though, in the stories told by younger women, the OW may be at odds with the OM, it is the OM who wins out. As they age, women still ac-

knowledge the relevance of nurturance and generativity, and still lodge these qualities in the figure of the OM, but they now propose that the more self-assertive, domineering OW will carry the day. Thus they imply that the soft, warm, maternal qualities are not so important on the family scene—or in their own affective repertoire—as when they were younger. Perhaps, as children become mature and less emotionally vulnerable, women can allow themselves to be more tolerant of their own needs for self-assertion and domination. Perhaps they are less frightened of the aggressive side of their personalities once they can reassure themselves that it will no longer interfere with the maternal function, will no longer have destructive consequences for the young.

Men. For male respondents, the age shift occurs along different continua than for women. Though for men, too, the OM's authority decreases with increasing age of respondent, the issues are those of competence and assertiveness. The 40-year-old male sees the OM as struggling with problems relating to assertion, guilt, nurturance, and affiliativeness—conflicts he attempts to solve in terms of complex role patterns that integrate the various elements. The 60-year-old male tends to see the OM as one who has simplified these problems. The OM has reached a solution through relinquishing the assertive role elements, and has abandoned any attempts at active manipulation of the environment. Rather than attempting to alter the environment, he adjusts to it. Rather than dealing with people and events, he organizes only the conceptual traces of the environment as they intrude upon his inner world of thought.

For older men, the figure of the OM

is elaborated largely in terms of his continuing cerebral competence (Category 6). Rather than actively intervening in the family and taking the responsibility such intervention entails, the OM restricts himself to the control of symbols. Ordering the inner cognitive world has replaced the more risk-laden transactions with the environment. For example, a man of 64 says of the OM: "He's easy-going. Doesn't care whether school keeps. He doesn't take care of his clothes." The theme of the story is illness, and the OM relative to this crisis is described further: "I don't know—but from his appearance he looks like a man who would like to light his pipe and get off somewhere and think about it." Here threat is met by flight and isolation; the OM's thought processes are the last remnant of competence, the surviving manipulative mode. The OM in this story has almost ceased dealing with a social environment. His only meaningful affiliation is to a nonhuman pacifier, a pipe. This, in rather extreme form, is the modal image of the OM given us by older men.

In relating these role descriptions with the personalities of respondents, the implication is that men, as they spell out the role of the OM, are concerned with the problems of male assertiveness. It will be recalled that somewhat the same issues are dealt with when men verbalize their fantasies around the YM. Men seem to say that, for both the YM and OM, aggression and self-assertion are inappropriate or unacceptable within the family. Young men must control themselves (or be controlled by parents). Older men, to avoid guilt, must act as altruists or conformists, or they must withdraw altogether from active participation in the family.

For both YM and OM, there is an area outside that of family interaction in which one can be competent—for the YM, the world of industry and work; for the OM, the world of thought.

With increasing age, men seem also to see a shifting pattern of interaction between YM and OM in the family scene. For the 40-year-old, a YM will direct his energies outward and the OM will be the benevolent authority in the family. For the 60-year-old, a YM has become a more assertive force within the family (although his assertiveness is still not harmful) and an OM has withdrawn from family interaction altogether. It is of special interest that the energies of the YM and OM do not clash; the two rarely strive for dominance in the same area at the same time.

Social-Class Differences

Table 4 shows great consistency between middle-class and working-class respondents of both sexes in the distribution of responses. It appears that social class is not a meaningful variable and that the same age and sex differences in perceptions of the OM occur in both social classes.

THE OLDER WOMAN

The Older Woman, by comparison to the other figures in the picture, is the key figure in the family. The family is her arena, and within it she emerges in full scope and complexity. In stories where the solidarity of the family is stressed, the OW is the one around whom the family is centered. In stories of conflict, she is always the key protagonist. There are, for instance, a number of stories that might be labeled "Oedipal" in theme—stories in which the YW is being claimed by the YM, or

vice-versa, and where there is conflict between the young and the old. In these stories, it is always the OW, but not the OM, who is seen as the protagonist in the struggle.

For the OW, the major issue is around retentiveness of the young—what is to be the extent and the nature of the tie between herself and her children. This issue always has strong emotional components for her.

In contrast to the Young Woman, the OW is seen as standing on her own feet, a person in her own right. The psychological distance between her and the Older Man is much greater than that between the two young figures. Whereas the YW and YM are seen in a close, collaborative relationship, the OW and OM stand separate and apart. They are often described in polar terms—if one is dominant, the other is submissive; if one is nurturant, the other is narcissistic.

Compared with the other figures, the OW is the one to whom the greatest depth and variety of feelings are assigned. In particular contrast to the YW, the OW is seldom bland and seldom neutral; she has strong feelings, both positive and negative. More often than not she is seen as acting on the basis of impulse and inner need; as the most inner-directed of the figures. Although her desires may be limited by the actions of others, or she may surrender to superior forces, she never doubts the rightness of her position.

The OW is not always seen as comfortable in her role, and she is the only figure who is as often described by respondents in negative as in positive terms.

Categories of Role Descriptions

Descriptions of the Older Woman were grouped into six major categories. While

the categories are ordered in Table 5 along the general continuum from submission to dominance, there are in reality two major themes which, in one or another combination, form the basis for the differentiation. The first is the theme of control over others—whether the OW is seen as submissive and controlled (Categories 1, 2, 3) or as dominant and the controller (Categories 4, 5, 6). The second theme is that of the nature of impulsivity—whether the OW is viewed as benign and nurturant (Categories 1, 4, 5) or self-assertive and aggressive (Categories 2, 3, 6).

1. *Submissive, nurturant.* Here the OW is viewed as passive, relative to the determination of outcomes. She is affiliative, nurturant, benevolent, but never self-assertive. She takes a position of deference to a wise, authoritative Older Man. She is the fluttery, "little woman" type and never intrudes on the masculine prerogatives of thought and decision. She is dependent on her husband for guidance and for control. To the extent that she takes action at all, the action is nurturant, promoting the best interests of others, especially the young.

2. *Controlled by OM.* Here the OW is seen as aggressive and impulsive, but she is controlled by the OM. Although she is something of a "battle-ax," she is more the family nuisance than the family menace. Her rages do not intimidate, they merely annoy. The wise and tolerant husband allows her free expression of her feelings, but deftly controls her. He determines outcomes and guarantees nurturant solutions to the autonomy-seeking young, often in the face of the OW's active opposition.

3. *Limited by children.* Here, as in the preceding category, an aggressive, domineering OW tries to extend her control over a resistant environment. While she now dominates her husband, she is successfully opposed and limited by the YW and/or the YM. The OM cannot provide a buffer between the intrusive OW and the young, but the young take up the cudgels for themselves and win out against the OW.

4. *The good mother.* Here the issue of dominance-submission is not specifically introduced, though the OW is implicitly given the decisive role in the disposition of affairs. The OW is the good, nurturant mother who guides and supports her gratified husband and children. She is mild, benign, maternal. Though she has

the most effective role in the family, there is no tension between her and the others. The view is rather of harmonious interaction, where it is only right and "natural" that the mother holds the most important place in the family.

5. *The matriarch.* In this category the OW is seen as a forceful and aggressive authority. While, however, she has complete sway over the others, this leads only to benign results. The family, rather than opposing her, bask contentedly in their dependent and submissive positions. Everyone benefits from her rule.

6. *Hostile self-assertion.* Here the OW is a stereotyped figure, one who exercises a harsh, arbitrary, and unopposed control. Her dominance is not tempered by any redeeming strain of affiliation or nurturance, nor does she have any concern for others. The OW is either pictured as the embodiment of amoral Id—all impulse and wrath; or the punitive Superego who harshly judges others and rigidly defines the moral code—a Superego armed, as it were, with the energies of the Id.

Age Differences

As shown in Table 5, role definitions of the Older Woman, like those of the other three figures, vary consistently with age of respondent. Whereas the age shift in the perception of the Older Man's role is in the direction of increasing submissiveness, the OW moves from a subordinate to an authoritative family role. (The age shift is statistically stable at the .001 level.)

In the first three categories, primarily those of younger respondents, the OW is seen either as "socialized"—adapting her behavior and attitudes to the needs of others (Category 1); or as aggressive and impulsive, but being held in check by

TABLE 5
ROLE DESCRIPTIONS OF THE OLDER WOMAN

Category:		Age 40-54		Age 55-70	
		Men	Women	Men	Women
1. Submissive, nurturant	Middle-class	10	6	1	3
	Working-class	4	4	2	3
	Total	24*		9	
2. Controlled by OM	Middle-class	1	9	1	0
	Working-class	5	5	1	2
	Total	20		4	
3. Limited by children	Middle-class	2	3	1	1
	Working-class	3	2	2	4
	Total	10		8	
4. The good mother	Middle-class	2	0	7	2
	Working-class	7	0	3	2
	Total	9		14	
5. The matriarch	Middle-class	1	3	0	3
	Working-class	1	0	1	1
	Total	5		5	
6. Hostile self-assertion	Middle-class	2	1	4	3
	Working-class	1	1	5	4
	Total	5		16	

* The chi-square test was applied to these category totals. The probability that the distribution occurred by chance is less than .001.

effective outer restraints (Categories 2, 3). In either case, her impact upon the situation is relatively unimportant. Older respondents hold that, whether "socialized" and nurturant (Categories 4, 5) or punitive and narcissistic (Category 6), the OW is the dominant figure, responsible for the outcome of family affairs.

Younger respondents, then, view the OW as sensitive to, or checked by, outer demands and pressures. Older respondents propose that the OW has come to be the embodiment of controls, strictures, limits. She has taken over the moral and directive qualities which, for younger respondents, were seen as operating outside herself.

In general, with increasing age of respondent, the OW emerges more and more as the most feelingful, demanding, and aggressive figure, as the other figures tend toward greater passivity, colorlessness, and conformity. In stories told by older respondents, the point at which the OW is described tends frequently to signal the break-through of impulsivity, as if the OW represents unchecked impulse in a scene otherwise populated by more constricted, conforming or affiliative figures.⁹

It has already been implied that, as regards respondents themselves, the Older Woman symbolizes the impulsive, self-centered qualities of the personality (in contrast to the OM who symbolizes ego qualities of the personality). The age shift in perception of the OW implies, therefore, increasing pressures from the impulse life in the face of decreasing ego controls.

⁹ Perhaps the projection of impulsive elements of personality upon the figure of the OW is partially stimulated by her "facelessness" in the picture. If impulsivity is regarded by respondents as "ego-alien," it might well be ascribed to that figure in the picture which provides fewest cues regarding social interaction.

Age-Sex Differences

Although men and women generally agree on the broad outlines of the shift from submission to dominance in the OW's role, the shift for the two sex groups takes place along different axes and in terms of different role attributes.

Men. While many men view the OW as having strong aggressive, intrusive needs, still a significantly greater number see her—whether dominant or not—as benign and maternal. (Sixty per cent of all males place the OW in Categories 1, 4, or 5.) The most frequent image of the OW among younger men is that of the good little woman leaning on a strong and wise husband (Category 1), or the good mother whose influence, though decisive, is only benevolent and constructive (Category 4). For older men, the most frequent image is again that of the good mother (Category 4) around whom there is no tension but only gratification.

For many of the men, then, the OW remains motherly, benign, nurturant, but the implication is that these maternal, affiliative qualities come to take on greater significance in the resolution of family issues. The OW remains essentially the same person, but her impact on the family increases. She is no longer submissive or controlled by others; instead, the issue of authority disappears altogether as the OW is allowed to "come into her own."

Perhaps the implication, as regards the personalities of most older men, is that they are more receptive to their own affiliative, nurturant, and sensual promptings than are younger men. In projecting these qualities upon the figure of the OW, and then giving the OW a central role in the determination of outcomes, these older men seem to be relatively acceptant of "womanly" qualities

in themselves, and to feel little need to deny or limit these qualities.

There are, on the other hand, a sizable number of older men who see the OW as a hostile and punitive autocrat (Category 6). In these stories, the OM withdraws from the family scene (removing himself, as it were, from the fray) but creates for himself an arena in which he is not only invulnerable but still competent, the arena of the intellect.

Perhaps this second group of older men, in contrast to the majority, cannot accept the new passivity in their personalities and so regress to archaic defenses against it. They project upon the Older Woman the responsibility for their passivity. By proposing that she is a tyrant, they justify their own submission as something imposed upon them, not something inherent in their own personalities.

This second group of older men are perhaps saying, also, that hostile and aggressive impulses in the personality (qualities projected upon the OW) can no longer, as with younger men, be countered with more benevolent and altruistic impulses. (It is only in the younger men's fantasies that the maternal OM limits and controls the impulsive OW.) Since impulse can now be only destructive and hostile, it should be split away altogether and the attempt should be to lead a life of quiet reason.

Women. While many women recognize the OW as a nurturant figure, the majority see her in terms of assertive, intrusive qualities. (55% of all female responses place the OW in Categories 2, 3, or 6.) Younger women tend to see the OW's needs for dominance and aggressive self-assertion as being effectively restrained—limits are set on her by her husband or by her children. (It should be noted that of 34 younger women re-

spondents, not one sees the OW in Category 4, as merely the "good mother" untroubled by problems of narcissistic needs.) For older women, the OW is essentially still a self-assertive figure, but now the limits have become ineffectual and less charged with moral authority. Where the OW was once blocked by agents who had "right" on their side—an OM who had greater wisdom and nurturance than she, or young people whose claims for autonomy were valid—the OW now becomes self-righteously assertive. This self-assertion may go in either direction—the good mother whose benevolent sway over family affairs is unchallenged, or the narcissistic autocrat—but in either case, it is the OW's domineering needs that win out.

Moving again to the area of personality: as has already been pointed out in reference to perceptions of the Older Man, women, as they age, seem to become more responsive toward, and less guilty about, their own aggressive, ego-centric impulses. Aggression and self-assertion are perhaps viewed as inimical to the central functions of motherhood, and can only be tolerated and acted out when the young are sufficiently independent to withstand or ignore the mother's intrusiveness. Hence it is only older women, whose children are presumably grown, who fantasy the OW's monolithic and unchecked dominance. The benign, maternal qualities tend meanwhile to become more ego-alien and are projected onto the OM, who, while seen as ineffectual, is nevertheless also regarded by women as the more lovable figure of the two.

It has been said that female respondents lay greater stress than male respondents on the OW's aggressive needs for controlling others and on the forces which oppose these needs. These findings

suggest a sex difference in regard to the more general issue of social controls. It appears that men, seeing the OW exercising a benign control over a complaisant and gratified family, are proposing generally that authority is not coercive, and that self-restriction and compliance are natural and unforced. Control is essentially internalized and does not require punitive external measures for its maintenance. Authority is good, acceptable, and unquestioned. This is in contrast to the female view. Women, viewing the domineering OW as essentially in conflict with her environment—forcibly controlling it, or being checked by it—suggest that for them control, submissiveness, and conformity are more difficult issues.

Social-Class Differences

Examination of Table 5 shows a remarkable consistency between middle-class and working-class groups in their views of the OW. As was the case with the figure of the Older Man, it is not social class, but age and sex of respondent that produce the variation in the data.

FANTASY DATA IN RELATION TO SEX-ROLE BEHAVIOR

Since these findings have been derived from projective data, what are their implications as regards role behavior?

The individual, in filling real life roles, resolves tensions between personal needs and social expectations. The task of the ego is to organize the various affective components of the personality into a personally expressive, though socially acceptable, pattern of behavior. When presented with the picture, however, a different demand is made of the respondent. He is asked, not to act in the real family setting, but to breathe vitality into a

representation of family life. The task of the ego is not one of integrating various aspects of the self into a coherent pattern of behavior, but the opposite: in effect, to distribute various components of the self among the various figures in the picture.

This fractionating of the components of personality makes the Thematic Apperception technique a valuable clinical instrument; but it imposes qualifications on its use in the study of social roles. In the latter case, the respondent describes a living complexity (the role of YM or OW) in terms of only one or a few facets of the self. The projected aspect of the self, temporarily winnowed out of the total personality structure, tends to be expressed in exaggerated form. The result is a certain stereotypy and a certain overemphasis in the role descriptions. The task for the role analyst is thus made correspondingly difficult. The role patterns he wishes to describe may have been distorted into nonviable extremes as they have become the foci for conflicting elements in the respondents' personalities. Rather than objective role descriptions, his data are the affective connotations of role behavior, those which people limit and modify in real life.

The findings presented here must be interpreted with caution, then, in applying them to actual role behavior. It should be kept in mind that if the respondent speaks of the Older Man as weak and passive, and the Older Woman as dominant and manipulative, he is not only describing two polar forms of behavior but also two aspects of himself, and that both such aspects will find some (though not equal) expression in his own behavior. If the respondent is an older man, he cannot be described merely on the basis of his description of the Older

Man, as passive and weak; for the respondent is a person who also has needs for strength and dominance (needs which may find release in the very behavior which otherwise implies passivity, such as hypochondria or "saintliness"). It is the nature of the task—responding to the picture—which allows him to describe the Older Man in more unitary ways than are actually true of himself.

These considerations apply equally with regard to collective role images that emerge from groups of respondents. For example, many older respondents seem to agree that older men are passive, affectless, and isolated from the ongoing stream of intrafamilial events. They are described as "smoking their pipes" and "thinking." It cannot be assumed, however, that the only role of older men in the family is to stand in the corner, thinking and smoking. (This is what is meant by a nonviable extreme.) People who live in the family setting, young or old, do interact with others and do impinge on the environments of other family members. What can be justifiably assumed from this image is not that older men never interact or relate, but that the very activities through which they express the outward forms of intimacy also tend to highlight their desire for passivity and isolated contemplation. The image does not report the daily reality of the older man's role; rather, it is a sharply drawn, condensed expression of the affective mode which underlies his activities. The sharpness of the image is derived from the condensed expression of what is seen as being central to the figure of the Older Man, and from the affective components of the respondents' personalities identified with this central tendency.

What we have in these data, then, is

centrality rather than experienced complexity of role behavior.

To take another example: in many stories, especially in those told by men, the description of the Older Woman provides a point at which unchecked impulse breaks into a scene otherwise peopled by more restrained or affiliative figures. She is a figure of primal omnipotence and wrath. "A devil. Very strict. Must run everything and everybody." In one sense the description functions to bring the aggressive impulse life of the respondent into the story.¹⁰ What emerges is not an unbiased account of the Older Woman, but a picture of the Older Woman as it is filled out by aggressive energy that has its locus within the respondent himself. (It is the respondent's own denied rage, projected onto the figure of the Older Woman, that he calls a "devil.") This account also represents a nonviable extreme. Women who live in a social environment cannot act purely from unmediated primitive impulse. They would soon be hospitalized, institutionalized, or dead. What we are being told in such accounts is that older women's behavior in the family expresses, for those respondents preoccupied with such issues, a central quality of free aggression.

Bearing such considerations in mind, these findings can nevertheless be related to actual role behavior. This relationship is posited on the grounds that the affective complexes energizing the perception of the stimulus figures are indeed cued

¹⁰ Just as in real life his wife may function so as to express elements of the respondent's impulse life that are denied expression in his own behavior. Our findings hint at the possibility that males often handle their aggression in the family by proposing that they are the passive object of attack from a woman, rather than by proposing that aggression stems from within themselves.

by the respondents' expectations of such figures in real life. Granted that various components of the respondent's own personality migrate toward one or another stimulus figure, the impressive fact is the consistency with which the same personality components migrate to the same figure in the picture. For instance, for both men and women respondents, it is almost always the Older Woman, and not the Older Man, to whom impulsivity, aggressivity, and hostile dominance are ascribed. This consistency cannot be explained by chance. The assumption seems warranted that there is something common to the actual role behaviors of older women that elicits this consistency in respondents' fantasies.

To sum up: projective data do not yield descriptions of the total and complex role of the older woman in the family as that role is expressed in everyday, overt behavior (similarly for other figures). What is obtained instead is a central aspect of the role; an aspect that, in one translated form of behavior or another, is being recognized by both men and women. To this extent, these findings reflect at least part of the social reality.

To this extent, furthermore, projective techniques such as the one used in this study can prove to be valuable instruments for understanding the underlying and persistent determinants of role behavior.

SUMMARY

The Thematic Apperception technique was used with a sample of middle-aged adults to study role-images of husbands, wives, sons, and daughters. A specially-designed picture showing a Young Man, a Young Woman, an Older Man, and an Older Woman was presented to 131 men and women aged 40

to 70 of two social-status levels, upper-middle class and working class.

The Young Man was generally seen as a somewhat detached figure in the family setting, oriented primarily toward the extrafamilial world, and related to parents in terms of formal rather than affective ties. The figure symbolized, for both men and women respondents, the primacy of social demands over personal needs for affiliation and self-expression.

The Young Woman was a bland figure, lacking in vitality or autonomy, but the figure around whom family issues were initiated and resolved. She provided the tie between younger and older generations. This figure symbolized issues related to tenderness, intimacy, and sexuality.

The Older Woman was seen as the key figure in the family, and the one who usually had the greatest impact upon the resolution of family issues. She was seen as struggling with problems of retaining and controlling the young, and with controlling her own needs for self-assertion. Compared with the other figures, the OW was ascribed the greatest variety and depth of feelings. The OW symbolized the impulsive, egocentric qualities of personality. Women respondents put stress upon the OW's aggressive qualities. Men saw her either as benign and maternal or as aggressively domineering.

The Older Man, by comparison with the OW, was a more neutral figure whose most consistent feelings were those of uncritical affiliation toward the others. He was seen as acting assertively only when such activity had the purpose of assisting the young in becoming autonomous. Otherwise, the OM was seen as conforming to the wishes of others or as withdrawn altogether from family interac-

tion. For both men and women respondents, the OM symbolized ego qualities of the personality (rationality, attention to the needs of others, competence) in relation both to environmental crises and to the pressures of the impulse life.

The role-images of all four figures varied consistently with age and sex of respondent, but not with social class. Most striking was the fact that with increasing age of respondents, the Older Man and Older Woman reversed roles in regard to authority in the family. For younger men and women (aged 40-54) the Older Man was seen as the authority figure. For older men and women (aged 55-70) the Older Woman was in the dominant role and the Older Man, no matter what other qualities were ascribed

to him, was seen as submissive.

The different images of all four figures presented by men and women at the two age levels are discussed in terms of implications for personality changes in the years from 40 to 70. For example, women, as they age, seem to become more tolerant of their own aggressive, egocentric impulses; while men, as they age, of their own nurturant and affiliative impulses. To take another example, with increasing age in both men and women, ego qualities in the personality seem to become more constricted—more detached from the mastery of affairs, and less in control of impulse life.

Discussed also are the advantages and disadvantages of projective data in understanding role behavior in real life.

REFERENCES

1. HENRY, W. E. The thematic apperception technique in the study of culture-personality relations. *Genet. Psychol. Monogr.*, 1947, 35, 3-315.
2. SCHAW, L. C., & HENRY, W. E. A method for the comparison of groups: A study in thematic apperception. *Genet. Psychol. Monogr.*, 1956, 54, 207-253.

(Accepted for publication November 5, 1957)





